



PCT

ORGANISATION MONDIALE DE LA PROPRIETE INTELLECTUELLE
Bureau international

DEMANDE INTERNATIONALE PUBLIEE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

| | | | |
|--|--|----|---|
| (51) Classification internationale des brevets ⁴ : C07K 7/10, 7/06, C12N 15/00 G01N 33/569, A61K 39/21, 37/02 | | A1 | (11) Numéro de publication internationale: WO 88/05440 (43) Date de publication internationale: 28 juillet 1988 (28.07.88) |
| (21) Numéro de la demande internationale: PCT/FR88/00025 (22) Date de dépôt international: 15 janvier 1988 (15.01.88) | | | (72) Inventeurs; et (75) Inventeurs/Déposants (<i>US seulement</i>): ALIZON, Marc [FR/FR]; 71, rue du Cardinal-Lemoine, F-75005 Paris (FR). MONTAGNIER, Luc [FR/FR]; 21, rue de Malabry, F-92350 Le-Plessis-Robinson (FR). GUETARD, Denise [FR/FR]; 4 B, rue Anselme-Payen, F-75015 Paris (FR). CLAVEL, François [FR/US]; 12103 Portree Drive, Rockville, MD 20852 (US). SONIGO, Pierre [FR/FR]; 23, rue Gutenberg, F-75015 Paris (FR). GUYADER, Mireille [FR/FR]; 68, rue Laugier, F-75017 Paris (FR). TIOLLAIS, Pierre [FR/FR]; 16, rue de la Glacière, F-75013 Paris (FR). CHAKRABARTI, Lisa [FR/FR]; 16, rue des 3 Portes, F-75005 Paris (FR). DESROSIERS, Ronald [US/US]; 13 Causeway Street, Udsom, MA 01749 (US). |
| (31) Numéros des demandes prioritaires: 003,764 87/01739 87/05398 | | | (74) Mandataires: GUTMANN Ernest etc.; S.C. Ernest Gutmann - Yves Plasserand, 67, boulevard Haussmann, F-75008 Paris (FR). |
| (32) Dates de priorité: 16 janvier 1987 (16.01.87) 11 février 1987 (11.02.87) 15 avril 1987 (15.04.87) | | | (81) Etats désignés: AU, DK, JP, KR, US. |
| (33) Pays de priorité: US FR FR | | | Publiée <i>Avec rapport de recherche internationale. Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si de telles modifications sont reçues.</i> |
| (60) Brevet ou demande principal(e) (63) Apparenté(e) par continuation US Déposée le 11 février 1987 (11.02.87) 013,477 (CIP) | | | |
| (71) Déposant (<i>pour tous les Etats désignés sauf US</i>): INSTITUT PASTEUR [FR/FR]; 25-28, rue du Dr.-Roux, F-75015 Paris (FR). . | | | |
| (54) Title: PEPTIDES HAVING IMMUNOLOGICAL PROPERTIES 2-HIV-2 (54) Titre: PEPTIDES AYANT DES PROPRIETES IMMUNOLOGIQUES 2-HIV-2 (57) Abstract Peptides having immunological properties in common with those of the peptidic skeleton of peptides of viruses of the family HIV-2, particularly the envelope glycoprotein of HIV-2, characterized in that they have also a peptidic structure in common with the peptidic skeleton of peptides of SIV, particularly the envelope glycoprotein of SIV. The invention also relates to diagnosis compositions capable of detecting an infection due to HIV-2 and to vaccine compositions. (57) Abrégé Peptides ayant des propriétés immunologiques en commun avec celles de l'ossature peptidique des peptides des virus de la classe HIV-2, notamment de la glycoprotéine d'enveloppe de HIV-2, caractérisés en ce qu'ils ont également une structure peptidique en commun avec l'ossature peptidique des peptides de SIV, notamment de la glycoprotéine d'enveloppe de SIV. L'invention concerne des compositions de diagnostic capable de détecter une infection due à HIV-2 et des compositions de vaccin. | | | |

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publant des demandes internationales en vertu du PCT.

| | | | | | |
|----|-----------------------------------|----|---|----|-----------------------|
| AT | Autriche | FR | France | ML | Mali |
| AU | Australie | GA | Gabon | MR | Mauritanie |
| BB | Barbade | GB | Royaume-Uni | MW | Malawi |
| BE | Belgique | HU | Hongrie | NL | Pays-Bas |
| BG | Bulgarie | IT | Italie | NO | Norvège |
| BJ | Bénin | JP | Japon | RO | Roumanie |
| BR | Brésil | KP | République populaire démocratique de Corée | SD | Soudan |
| CF | République Centrafricaine | KR | République de Corée | SE | Suède |
| CG | Congo | LI | Liechtenstein | SN | Sénégal |
| CH | Suisse | LK | Sri Lanka | SU | Union soviétique |
| CM | Cameroon | LU | Luxembourg | TD | Tchad |
| DE | Allemagne, République fédérale d' | MC | Monaco | TG | Togo |
| DK | Danemark | MG | Madagascar | US | Etats-Unis d'Amérique |
| FI | Finlande | | | | |

5

Peptides ayant des propriétés immunologiques de HIV-2

La présente invention est relative à des peptides ayant des propriétés immunologiques, le cas échéant immunogènes, en commun avec des antigènes susceptibles d'être obtenus sous une forme purifiée, à partir de virus capables de provoquer des lymphadénopathies susceptibles d'être relayées ensuite par le syndrome d'immunodéficience acquise (SIDA) chez l'homme.

L'invention concerne en particulier des peptides antigéniques susceptibles d'être reconnus par des anticorps induits chez l'homme par des virus désignés par l'abréviation HIV, selon la nomenclature définie dans NATURE. Elle concerne également des peptides ayant des propriétés immunogènes ou susceptibles d'être rendus immunogènes in vivo, cette immunogénicité étant susceptible de se manifester par l'induction in vivo d'anticorps reconnaissant des antigènes caractéristiques des virus HIV-2 et même, au moins en ce qui concerne certains de ces peptides, des antigènes issus de HIV-1.

L'invention concerne en outre des applications de ces peptides à la fabrication de compositions pour le diagnostic in vitro chez l'homme de potentialité de certaines formes du SIDA et, en ce qui concerne certains d'entre eux, à la production de compositions immunogènes et de compositions vaccinantes contre les rétrovirus HIV.

De même l'invention concerne les applications aux mêmes fins des anticorps susceptibles d'être induits

in vivo par les peptides immunogènes ou rendus immuno-gènes et, pour certains de ces anticorps, leurs applica-tions à la production de principes actifs de médicaments contre ces SIDAS humains.

5 L'invention concerne également la mise en oeuvre de certains de ces peptides dans des procédés pour le diagnostic in vitro chez l'homme de certaines formes du SIDA, ainsi que leur application à la consti-tution de trousse ou "kits" de diagnostic.

10 Un premier rétrovirus dénommé LAV-1 ou HIV-1 a été isolé et décrit dans la demande de brevet GB.83/24.800 et une demande EP.84/401.834 du 14/09/84. Ce virus a également été décrit par F.Barre Sinoussi et al. dans Science, 220 n° 45-99, 20 pages 868-871.

15 Des variants de ce virus HIV-1 désignés par LAV ELI et LAV MAL, ont également été isolés, caracté-risés et décrits dans la demande de brevet EP.84/-401.834.

20 Les virus HIV-1 et leurs variants possèdent les propriétés suivantes :

- ils ont pour cibles préférentielles les cellules Leu3 (ou lymphocytes T4) humaines et leurs cellules dérivées "immortalisées".

25 - ils ont une activité transcriptase inverse nécessitant la présence d'ions Mg²⁺ et présentent une forte activité pour le poly(adenylate-oligo-deoxythymi-dylase) poly(A)-oligo(dT)12-18)

- ils ont une densité de 1,16 à 1,17 sur gradient de sucre,

30 - ils ont un diamètre moyen de 139 nanomètres et un noyau de diamètre moyen de 41 nanomètres,

- les lysats de ces virus contiennent une protéine p25 (protéine du noyau) qui ne croise pas im-munologiquement avec la protéine p24 de HTLV-1,

35 - ils contiennent une protéine p42 appartenant à leur enveloppe,

- ils contiennent également une glycoprotéine d'enveloppe gp110 d'un poids moléculaire de 110.000.

L'isolement et la caractérisation de rétro-virus appartenant à une classe distincte et n'ayant qu'une parenté immunologique réduite avec les précédents, ont été décrits dans la demande de brevet européen n° 87/400.151.4. Ces rétrovirus qui ont été regroupés sous la désignation HIV-2, ont été isolés chez plusieurs malades africains présentant des symptômes d'une lymphadénopathie ou d'un SIDA.

Les rétrovirus du type HIV-2 comme les rétro-virus du type HIV-1, se caractérisent par un tropisme pour les lymphocytes T4 humains et par un effet cytopathogène à l'égard de ces lymphocytes, lorsqu'ils s'y multiplient, pour alors causer soit des poly-adénopathies généralisées et persistantes, soit un SIDA.

Plus généralement les rétrovirus purifiés par HIV-2 possèdent en général les propriétés suivantes :

- la cible préférentielle des rétrovirus HIV-2 est constituée par les cellules Leu3 (ou lymphocytes T4) humaines et pour des cellules "immortalisées" dérivées de ces lymphocytes T4 ;
- ils sont cytotoxiques pour les lymphocytes T4 humains
- ils ont une activité de transcriptase inverse nécessitant la présence d'ions Mg^{2+} et présentant une forte activité pour le poly(adénylate-oligodéoxythymidylase) (poly(A)-oligo(dT) 12-18) ;
- ils ont une densité de 1,16 dans un gradient de sucreose ;
- ils ont un diamètre moyen de 140 nanomètres et un noyau ayant un diamètre moyen de 41 nanomètres ;
- ils peuvent être cultivés dans des lignées permanentes du type HUT ou exprimant la protéine T4 ;
- ils ne sont pas infectieux pour les lymphocytes T8 ;
- les lysats de ces virus contiennent une protéine p26

- qui ne croise pas immunologiquement avec la protéine p24 du virus HTLV-I ou du virus HTLV-II ;
- ces lysats contiennent en outre une protéine p16 qui n'est pas reconnue immunologiquement par la protéine p19 de HTLV-I ou de HTLV-II dans des essais de radioimmuno-précipitation ;
 - ils contiennent en outre une glycoprotéine d'enveloppe ayant un poids moléculaire de l'ordre de 130.000-140.000 qui ne croise pas immunologiquement avec la gp110 des HIV-1, mais qui en revanche croise immunologiquement avec la glycoprotéine d'enveloppe gp140 de STLV-III (virus isolé chez le singe) ;
 - ces lysats contiennent encore des antigènes marquables par la ³⁵S-cystéine, dont les poids moléculaires s'étagent entre 32.000 et 42.000-45.000 : ils comprennent notamment un antigène ayant un poids moléculaire de l'ordre de 36.000 et un antigène ayant un poids moléculaire de l'ordre de 42.000, l'un de ces antigènes (p36 et p42) constituant vraisemblablement une glycoprotéine trans-membranaire du virus HIV-2 ;
 - l'ARN génomique des HIV-2 n'hybride pas avec l'ARN génomique de HIV-1 dans des conditions stringentes ;
 - dans des conditions non stringentes, l'ARN génomique de HIV-2 n'hybride, ni avec le gène env et le LTR qui le jouxte, de HIV-1, ni avec des séquences de la région pol du génome de HIV-1 ;
 - dans des conditions non stringentes, il hybride faiblement avec des séquences de nucléotides de la région de HIV-1.

Un autre rétrovirus dénommé SIV-1, cette dénomination remplaçant la dénomination antérieurement connue STLV III, a été isolé chez le singe macaque rhésus. (M.D.Daniel et al. Science 228, 1201 (1985) N.L.Letwin et al, Science 230, 71 (1985) sous l'appellation "STLV-III^{mac}").

Un autre rétrovirus, désigné "STLV-III_{AGM}", (ou SIV_{AGM}) a été isolé chez des singes verts sauvages. Mais, contrairement au virus présent chez le singe macaque rhésus, la présence de "STLV-III_{AGM}" ne semble pas induire une maladie du type SIDA chez le singe vert d'Afrique.

Une souche du rétrovirus SIV-1mac a été déposée à la CNCM le 7 Février 1986 sous le n° I-521. Des études ont montré que le rétrovirus SIV-1 comporte certaines protéines possédant une certaine parenté immunologique avec des protéines ou glycoprotéines structurales susceptibles d'être obtenues dans des conditions analogues, à partir de HIV-2. Ce rétrovirus SIV-1, dont on a constaté le caractère infectieux chez les singes, avait été désigné par STL VIII par les chercheurs qui l'ont isolé (références bibliographiques précitées).

Pour la commodité du langage, ces virus ne seront plus désignés dans ce qui suit que par l'expression SIV (l'expression SIV est l'abréviation anglaise de "Simian Immunodeficiency Virus" (virus d'immunodéficience du singe)) éventuellement suivi d'une abréviation désignant l'espèce de singe dont ils sont issus, par exemple, MAC (ou mac) pour le macaque ou AGM pour le singe vert d'Afrique (abréviation de "African Green Monkey").

En mettant en oeuvre les mêmes techniques que celles rappelées plus haut, il a été constaté que l'on pouvait également obtenir à partir de SIV-1mac :

- une protéine principale du noyau p27, ayant un poids moléculaire de l'ordre de 27 kilodaltons,
- une glycoprotéine majeure d'enveloppe, gp140,
- une protéine vraisemblablement transmembranaire p32, qui n'est guère observée en RIPA lorsque le virus a au préalable été marqué par la ³⁵S-cystéine, mais qui peut

être observée dans les essais d'immunoempreintes (Western blots), sous forme de bandes larges.

Des études plus précises ont été réalisées en ce qui concerne les précédents virus HIV-2 et SIV. La poursuite de l'étude des rétrovirus HIV-2 a également conduit à l'obtention de séquences d'ADN complémentaires (ADNc) des ARNs de leurs génomes. La séquence nucléotidique complète de l'ADNc d'un rétrovirus représentatif de la classe HIV-2 (HIV-2 ROD) a été déposée le 21/02/-10 1986 à la CNCM sous le n° I-522, sous le nom de référence LAV-II ROD).

Cette séquence nucléotidique et les phases de lecture ouverte qu'elle contient sont indiqués à la figure 1 A.

En outre, la poursuite de l'étude d'autres rétrovirus a également permis d'aboutir à l'obtention de leurs séquences nucléotidiques complètes. Il en est en particulier ainsi de l'ADNc dérivé de l'ARN génomique de SIV.

Le clonage et le séquençage du virus SIV-1mac qui ont permis l'obtention de sa séquence nucléotidique ont été réalisés dans les conditions suivantes :

L'ADN de cellules HUT 78 infectées par le virus SIV (isolat STLV-III mac 142-83 décrit par Daniel et al. (1985) Science, 228, p.1201-1204, digéré partiellement par l'enzyme de restriction Sau3A a été cloné au site BamHI du bactériophage vecteur Lambda EBL3 pour constituer une banque génomique. Les 2 millions de phages recombinants de la banque génomique ainsi constituée ont été criblés in situ en conditions de sécurité P3, à l'aide de séquences du virus HIV2 provenant des clones lambda-ROD4, lambda-ROD35 et E2 (Clavel et al. (1986-Nature, 324, p.691.) et nick-translatées.

L'hybridation a été réalisée en 5xSSC à 50°C et les lavages en 2xSSC à 50°C. Un seul clone contenant

l'ensemble des séquences virales a été obtenu. Ce clone est désigné par lambda-SIV-1. L'insérat du phage lambda-SIV-1 mesure 16,5 kb au total et comprend un provirus intégré auquel manquent seulement les 250 premières bases du LTR gauche, alors que le LTR droit est complet.

Le provirus intégré a été séquencé par la méthode des didéoxynucléotides après sous-clonage de fragments aléatoires dans le phage M13mp8. 300 sous-clones ont été analysés.

Des fragments d'ADNc provenant du clone Lambda SIV-1 insérés dans des plasmides pSIV-1.1 et pSIV-1.2 ont été déposés à la CNCM le 15 Avril 1987, sous les numéros I-658 (pSIV-1.1) et I-659 (pSIV-1.2).

Les résultats ont été mentionnés dans les figures décrites ci-après.

La figure 1B représente la séquence nucléotidique du génome viral de SIV et les séquences qui en sont déduites pour les protéines virales correspondant aux produits des gènes gag, pol, env, Q, X, R, tat, art, F.

Les figures 3 à 11 et la figure 1C représentent les comparaisons des produits théoriques des gènes viraux et des LTR entre HIV2 et SIVmac. (λ SIV-1).

L'invention concerne de plus les fragments d'ADNc déduits de l'ADNc issu du génome entier de SIV-1, ces fragments contenant une ou plusieurs séquences issues de la séquence complète d'ADNc et qui codent pour des peptides intéressants de l'invention. Ces séquences sont indiquées à la figure 1B et, à la figure 1C pour ce qui a trait à la séquence LTR du virus,

Les séquences nucléotidiques de l'ADNc de SIV ont été placées en correspondance avec les séquences nucléotidiques du virus HIV-2 ROD pour ce qui concerne la séquence LTR (figure 1C). Cette présentation que l'on retrouve pour le génome entier en rapprochant la figure 1B

des figures 3 à 11 permet de repérer ou de déduire les acides nucléiques ayant des éléments de structure essentiels communs aux deux virus.

L'invention concerne naturellement aussi l'utilisation des cADNs issus de SIV ou de leurs fragments (ou de recombinants les contenant) en tant que sondes, pour le diagnostic de la présence ou non de virus HIV-2 dans des échantillons de sérum ou d'autres liquides ou tissus biologiques obtenus à partir de patients suspectés d'être porteurs du virus HIV-2. Ces sondes sont de préférence marquées également (marqueurs radio-actifs, enzymatiques, fluorescents, etc.). Des sondes particulièrement intéressantes pour la mise en oeuvre du procédé de diagnostic du virus HIV-2 ou d'un variant de HIV-2 peuvent être caractérisées en ce qu'elles comprennent la totalité ou une fraction de l'ADNc complémentaire du génome du virus SIV ou encore notamment les fragments recombinants contenus dans divers clones.

Les sondes mises en oeuvre dans ce procédé de diagnostic du virus HIV-2 et dans les kits de diagnostic ne sont en aucune façon réduites aux sondes décrites précédemment. Elles comprennent au contraire toutes les séquences nucléotidiques issues du génome du virus SIV, d'un variant de SIV ou d'un virus proche par sa structure, dès lors qu'elles permettent la détection dans des fluides biologiques de personnes susceptibles de développer un SIDA, d'anticorps dirigés contre un HIV-2 ou d'un virus qui en est proche.

La détection peut être réalisée de toutes façons en soi connues. Elle peut comprendre une mise en contact de ces sondes soit avec les acides nucléiques obtenus à partir des cellules contenues dans ces sérum ou autres milieux biologiques, par exemple liquides céphalo-rachidiens, salives, etc... Elle peut aussi

comprendre une mise en contact de ces sondes avec ces milieux eux-mêmes dès lors que leurs acides nucléiques ont été rendus accessibles à l'hybridation avec ces sondes, et ce dans des conditions permettant l'hybridation entre ces sondes et ces acides nucléiques. L'é-
5 tape finale du diagnostic in vitro comprend alors la détection de l'hybridation éventuellement produite. Le susdit diagnostic mettant en jeu des réactions d'hy-
bridation peut également être réalisé à l'aide de mé-
10 langes de sondes respectivement originaires d'un HIV-2 et d'un SIV-1 ou d'un HIV-1, d'un HIV-2 et d'un SIV, dès lors qu'il n'est pas nécessaire de faire une différence entre le type de virus recherché.

D'une façon générale, le procédé de diagnostic
15 de la présence ou non du virus HIV-2 ou d'un variant dans des échantillons de sérums ou d'autres liquides ou tissus obtenus à partir de patients suspectés d'être porteurs du virus HIV-2 comprend les étapes suivantes :

1/ au moins une étape d'hybridation conduite
20 dans des conditions stringentes, par mise en contact de l'ADN de cellules de l'échantillon du patient suspect avec l'une des susdites sondes marquées sur une membrane appropriée,

2/ le lavage de ladite membrane avec une solution assurant la conservation de ces conditions strin-
25 gentes de l'hybridation,

3/ la détection de la présence ou non du virus HIV-2 par une méthode d'immunodétection.

Dans un autre mode de réalisation préféré du
30 procédé selon l'invention l'hybridation précitée est conduite dans des conditions non stringentes et le la-
vage de la membrane est réalisé dans des conditions adaptées à celles de l'hybridation.

Il va de soi que l'invention concerne les
35 acides nucléiques correspondant à des séquences placées

en des régions analogues de variants de SIV ainsi que tous les acides nucléiques dont les modifications résulteraient de la mise à profit de la dégénérescence du code génétique.

5 Les études comparatives qui ont aussi permis d'aboutir à des résultats relatifs aux protéines de noyau (core), ci-après dénommées "protéines gag" et aux protéines d'enveloppes, ci-après dénommées "protéines env", ont également été rapportés dans la demande de brevet européen n° 87/400.151.4, déjà citée. Ces résultats montrent que les protéines du noyau (protéines gag) dans HIV-2 présentent des différences moins accentuées par rapport à celles des virus HIV-1, que les protéines d'enveloppe (protéines env). Globalement les protéines env dans HIV-2 se sont révélées présenter des parentés immunologiques extrêmement faibles, sinon inexistantes, avec les protéines env correspondantes des virus HIV-1.

20 Au contraire des études comparatives effectuées entre les structures des séquences d'ADNc des virus HIV-2 et SIV permettent de mettre en évidence certaines caractéristiques communes qui apparaissent au niveau des protéines.

25 Globalement, les protéines de HIV-2 et de SIV-1 montrent des parentés immunologiques importantes.

La glycoprotéine majeure d'enveloppe de HIV-2 s'est révélée être plus proche immunologiquement de la glycoprotéine majeure d'enveloppe de SIV que de la glycoprotéine majeure d'enveloppe de HIV-1.

30 Ces constatations s'imposent non seulement au niveau des poids moléculaires : 130-140 kilodaltons pour les glycoprotéines majeures de HIV-2 et de SIV contre environ 110 pour la glycoprotéine majeure d'enveloppe de HIV-1, mais aussi au niveau des propriétés immunologiques, puisque des sérums prélevés à partir de malades

infectés par HIV-2, et plus particulièrement des anticorps formés contre la gp140 de HIV-2 reconnaissent la gp140 de SIV-1mac, alors que dans des essais semblables les mêmes sérum et les mêmes anticorps de HIV-2 ne reconnaissent pas la gp110 de HIV-1. Mais les sérum anti-HIV-1 qui n'ont jamais réagi avec la gp140 de HIV-2 précipitent une protéine de 26 Kdal marquée par la ³⁵S-cystéine, contenue dans les extraits de HIV-2.

La protéine majeure du noyau (core) de HIV-2 semble présenter un poids moléculaire moyen (environ 10 26.000) intermédiaire entre celui de la p25 de HIV-1 et la p27 de SIV.

Ces observations résultent des essais réalisés avec des extraits viraux obtenus à partir du HIV-2 isolé 15 à partir de l'un des patients susmentionnés. Des résultats similaires ont été obtenus avec des extraits viraux du HIV-2 isolé à partir du second patient.

Des études plus poussées ont conduit les inventeurs à reconnaître une première classe de peptides 20 ayant des séquences d'aminoacides soit identiques, soit proches de séquences contenues à l'intérieur des structures des protéines gag et env de HIV-2 ou de SIV voire de HIV-1. Ces peptides sont notamment applicables au diagnostic d'une infection chez l'homme par le virus HIV-2 ou de l'un de ses variants.

A cet égard la présente invention concerne également des procédés et des compositions de diagnostic pour la détection in vitro d'anticorps dirigés contre un virus HIV-2 ou de ses variants, plus particulièrement dans des échantillons biologiques, notamment des sérum de patients ayant subi une infection par le virus HIV-2, certains de ces peptides permettant une discrimination particulièrement poussée entre les infections dues à des virus HIV-2 et à des virus HIV-1.

35 Ces études poussées ont également conduit à la

possibilité de synthétiser des peptides immunogènes ou susceptibles d'être rendus immunogènes, présentant des caractéristiques de structures leur permettant d'induire *in vivo* la production d'anticorps susceptibles de reconnaître des protéines env à la fois dans HIV-1 et dans HIV-2 et, au moins pour certains de ces peptides, de se fixer tant sur des virus HIV-1 que sur des virus HIV-2, plus particulièrement aux fins de les neutraliser. L'utilisation de ces derniers types de peptides est donc particulièrement indiquée pour la production de principes actifs de vaccins contre les virus HIV, donc contre le SIDA.

Pour désigner ci-après les résidus d'amino-acides entrant dans la constitution des peptides selon l'invention, on aura recours, pour ceux des acides aminés ayant une signification univoque à la nomenclature internationale désignant chaque acide aminé naturel par une lettre unique (lettre majuscule) selon le tableau des correspondances qui suit :

| | | |
|----|---|--------------------|
| 20 | M | Méthionine |
| | L | Leucine |
| | I | Isoleucine |
| | V | Valine |
| | F | Phénylalanine |
| 25 | S | Sérine |
| | P | Proline |
| | T | Thréonine |
| | A | Alanine |
| | Y | Tyrosine |
| 30 | H | Histidine |
| | Q | Glutamine |
| | N | Asparagine |
| | K | Lysine |
| | D | Acide Aspartique |
| 35 | E | Acide glutaminique |

| | |
|---|-------------|
| C | Cystéine |
| W | Tryptophane |
| R | Arginine |
| G | Glycine |

5 Lorsqu'un acide aminé pourra, en raison de sa position au sein de la chaîne d'aminoacides caractéristique d'un peptide déterminé, prendre plusieurs significations, il pourra soit être désigné par un tiret "--", si sa signification peut être quelconque, soit par une lettre minuscule lorsque cet aminoacide pourra 10 présenter un nombre limité de significations préférées, ce nombre étant cependant toujours supérieur à 1. Dans ce dernier cas, les significations possibles de cette lettre minuscule seront toujours précisées en rapport 15 avec le peptide auquel il appartient.

15 Afin de faciliter la lecture, ces peptides seront désignés par une abréviation env ou gag suivie 20 d'un indice numérique, par référence à des séquences d'aminoacides contenues, selon le cas, soit dans les protéines env soit dans les protéines gag de certains HIV-1, HIV-2 ou SIV. Il y sera encore fait référence dans ce qui suit.

Enfin dans les définitions qui suivent

- les groupes X représentent soit un groupe NH₂ libre ou amidé, notamment par un ou deux groupes alcoyle comprenant de 1 à 5 atomes de carbone, soit un groupe peptidique comprenant de 1 à 5 aminoacides, dont l'aminoacide N-terminal présente lui-même un groupe NH₂ libre ou amidé comme précédemment indiqué, et
- les groupes Z représentent, soit un groupe -OH libre ou alcoyle et contenant alors un groupe alcoyle comprenant de 1 à 5 atomes de carbone, soit un groupe peptidique comprenant de 1 à 5 aminoacides, dont l'aminoacide C-terminal présente lui-même un groupe -OH libre ou alcoyle, comme précédemment indiqué, les

groupes de 1 à 5 acides aminés le cas échéant contenus dans X ou Z ou dans les deux à la fois étant tels, que leur présence n'est pas incompatible avec la préservation pour l'essentiel des propriétés immunologiques, le cas échéant immunogènes, des peptides qui en sont dépourvus.

Les peptides selon l'invention, qui ont en commun des propriétés immunologiques avec des antigènes de HIV-2 et, pour certains d'entre eux également avec des antigènes de HIV-1 ou de ses variants, sont caractérisés en ce qu'ils ont également une structure peptidique en commun avec les antigènes de SIV. De façon avantageuse, ces peptides comprennent normalement au plus 40 résidus d'acides aminés.

Des peptides préférés sont les suivants :

15 env1

XRV-AIEKYL-DQA-LN-WGCAFRQVCZ

env2

X-LE-AQI-QQEKNMYELQKLNZ

20 env3

XELG DYKLVEITPIG-APT--KR-----Z

env4

X----VTV-YGVP-WK-AT--LFCA-Z

env5

X---QE--L-NVTE-F--W-NZ

25 env6

XL---S-KPCVKLTPLCV--Z

env7

X---N-S-IT--C-K----Z

30 env8

X-I---YC-P-G-A-L-C-N-TZ

env9

X-----A-C-----W--Z

env10

35 X-G-DPE-----NC-GEF-YCN-----NZ

15

env11

X-----C-IKQ-I-----G---YZ

Plus particulièrement l'invention concerne les peptides suivants :

5 env1

XRV-AIEKYL-DQA-LN-WGCAFRQVCZ

env2

X-LE-AQIQQEKNMYELQKLNZ

env3

10 XELGDYKLVEITPIG-APT--KR----Z

env4

X----VTV-YGVP-W--AT--LFCA-Z

env5

X----E--L-NVTE-F--W-NZ

15 env6

XL---S-KPCVKL-PLC---Z

env7

X---N-S-I---C-K----Z

env8

20 X-I---YC-P-G-A-L-C-N-TZ

env9

X-----A-C-----W--Z

env10

X-G-DPE-----NC-GEF-YC-----NZ

25 env11

X-----C-I-Q-I-----G---YZ

Des peptides avantageux correspondant aux précédents, présentent les formules qui suivent :

env1

30 XRVTAIEKYLQDQARLNSWGCAFRQVCZ, ou

XRVTAIEKYLKDQALNQAWGCAFRQVCZ

env2

XSLEQAQIQQEKNMYELQKLNWZ, ou

XLLEEAQIQQEKNMYELQKLNWZ

env3

XELGDYKLVEITPIGFAPTKEKRYSSAHZ, ou

XELGDYKLVEITPIGLAPTNVKRYTTG-Z

5 (On remarquera que les peptides env1, env2, env3 attestent de la très grande parenté entre HIV-2 et SIV-1. En effet le premier peptide est inclu dans le génome de HIV-2 et le second, dans celui de SIV-1).

env4

XabcdVTVeYGV PfWogATHiLFCAjZ,

10 dans lesquels les lettres de a à j peuvent avoir les significations suivantes :

a est C, E ou D

b est T, K, D, N ou I

c est Q ou L

d est Y ou W

15 e est F ou Y

f est T, V ou A

g est N ou E

h est I ou T

20 i est P ou T

j est T ou S

o est K ou R

env5

XabcoEdeLfNVTEgFhiWjNZ,

25 dans lequel les lettres de a à j peuvent avoir les significations suivantes :

a est D ou P

b est D ou N

c est Y ou P

30 d est I, V, I ou L

e est T, V, E ou A

f est V, G ou E ou -

g est A, N, G ou S

h est D ou N

35 i est A ou M

j est N, K ou E

o est Q ou S

env6

XLabcsdKPCVKLoPLCuefKZ,

5 dans lequel les lettres de a à f peuvent avoir les significations suivantes :

a est F ou W

b est E ou D

c est T ou Q

10 d est I ou L

e est A, S ou T .

f est M ou L

o est T ou S

u est V ou I

env7

XabCNxSyIocdCeKfghiz,

dans lequel les lettres de a à i et x et y peuvent avoir les significations suivantes :

a est N ou T ou I

20 b est H ou S ou N

c est E ou Q

d est S, A ou C

e est D ou P

f est H, V ou D

25 g est Y ou S

h est W ou F

i est D ou E

x est T ou R

y est V ou A

30 o est T ou Q

env8

XaIbcdYCxPeGfAgLhCiNjTz,

dans lequel les lettres de a à k et x peuvent avoir les significations suivantes :

a est A ou P
b est R ou P
c est F, I ou C
d est R ou H
e est P ou A
5 f est Y ou F
g est L ou I
h est R ou K
i est - ou N
j est D ou K
10 x est A ou T

env9

XwabcxyAdCefghizWjkZ,
dans lequel les lettres de a à k et x à z peuvent avoir
les significations suivantes :

15 a est K ou - ou E
b est R ou -
c est P ou M ou I
d est W ou H ou Y
e est W ou N ou T ou R
20 f est F ou I
g est K ou S ou N ou G
h est G ou R ou E
i est - ou A ou T
j est K ou N ou D ou S

25 k est D ou A ou N ou K ou E
w est N, D ou I
x est R ou G ou K
y est Q ou K ou R
z est K ou E ou Q ou N
30 env10

XaGbDPEcdefghNCiGEFjYCokxlmnNZ,
dans lequel les lettres de a à n et x peuvent avoir les
significations suivantes :

a est K ou - ou G
b est S ou G ou -
c est V ou I
d est A ou V ou T
e est Y ou T ou M ou F
5 f est M ou H
g est W ou S
h est T ou F
i est R ou G
j est L ou F
10 o est N ou K
k est M ou S
l est W ou Q ou K ou G
m est F ou L
n est L ou F
15 x est T ou S ou N

env11

XabcdwCeIoQfIxgyhizGjklyZ,
dans lequel les lettres de a à l et w à z peuvent avoir
les significations suivantes :
20 a est R ou T ou S ou N
b est N ou I
c est Y ou T
d est A ou L ou V
e est H ou R
25 f est I ou F
g est T ou M
h est H ou Q ou A
i est K ou E
j est R ou K
30 k est N ou A
l est V ou M
w est P ou Q
x est N ou K
y est W ou V
35

z est V ou T ou K

o est K ou R

La structure du peptide antigénique codé par le gène gag et désigné par gag1 est également représentée ci-après :

5 XDCKLVLKGLGaNPTLEEMLTAZ,

dans lequel la lettre a désigne M ou T.

10 Il sera remarqué que, d'une façon générale, les aminoacides ayant une signification univoque (donc représentés par une lettre majuscule correspondant à la nomenclature internationale) qui interviennent dans les définitions qui précèdent des peptides selon l'invention, se trouvent être la correspondance avec des aminoacides identiques placés dans le même ordre dans les séquences env ou gag correspondantes de la protéine env ou 15 gag d'au moins l'un des HIV, ou de SIV-1.

20 Les positions de ces séquences sont soulignées et repérées au sein des séquences d'aminoacides des protéines env respectivement de HIV-2 ROD (CNCM n° I-532) et HIV-1 BRU (CNCM n° I-232) représentées à la figure 2. Par ailleurs, les alignements des acides aminés des protéines env et gag respectivement de SIV-1mac (CNCM n° I.521) et de HIV-2 ROD sont présentées à la figure 3 et à la figure 4.

25 Les traits pleins qui apparaissent en certaines localisations de ces séquences visent à souligner que certains aminoacides contenus dans ces séquences ont été volontairement déletés au plan de la présentation, afin de permettre la mise en alignement d'aminoacides respectivement identiques (alors marqués d'un astérisque) ou de deux points verticaux sur une même ligne 30 verticale dans les séquences des protéines correspondantes de HIV-1 et de HIV-2 d'une part, de SIV et de HIV-2 d'autre part.

Outre les peptides précités, l'invention concerne également les peptides modifiés par insertion et/ou délétion et/ou substitution d'un ou plusieurs acides aminés, pour autant que les propriétés antigéniques ou immunogènes desdits peptides ne sont pas modifiées, ou que les propriétés de reconnaissance de l'antigène ou de l'anticorps avec lesdits peptides ne sont pas substantiellement modifiées.

Dans un mode de réalisation particulièrement préféré, l'invention concerne des peptides ayant des propriétés immunologiques en commun avec l'ossature peptidique de la glycoprotéine d'enveloppe des virus de la classe HIV-2, ces peptides contenant un nombre de résidus d'acides aminés n'excédant pas 40.

Ces peptides préférés selon l'invention ont les séquences suivantes :

env1

RVTAIEKYLQDQARLNNSWGCAFRQVC

AIEKYLQDQ

RVSAIEKYLKDQAQLNAWGCAFRQVC

AIEKYLKDQ

env2

SLEQAAQIQQEKNMYELQKLNSW

QIQQEKN

LLEEAQIQQEKNMYELQKLNSW

env3

ELGDYKLVEITPIGFAPTKEKRYSSAH

YKLVEITPIGFAPTKEK

ELGDYKLVEITPIGLAPTNVKRYTTG-

YKLVEITPIGLAPTNVK

env4

CTQYVTVFYGVPTWKNATIPLFCAT

VTVFYGVPTWKNAT

CIQYVTVFYGVPAWRNATIPLFCAT

VTVFYGVPAWRNAT

EKLWVTVYYGVPVWKEATTLFCAS
VTVYYGVPVWKEAT
EDLWVTVYYGVPVWKEATTLFCAS.
VTVYYGVPVWKEAT
DNLWVTVYYGVPVWKEATTLFCAS
5 VTIVYYGVPVWKEAT
env5
DDYQEITL-NVTEAFDAWN
L-NVTEAF
DDYSELAL-NVTESFDAWEN
10 L-NVTESF
PNPQEUVVLVNVTENFNMWKN
LVNVTENF
PNPQEIELENVTEGFNMWKN
LENVTEGF
15 PNPQEIALENVTEFNFMWKN
LENVTENF
env6
ETSIKPCVKLTPLCVAMK
ETSIKPCVKLSPLCITMR
20 DQSLKPCVKLTPLCVSLK
DQSLKPCVKLTPLCVTLN
PCVKLTPLCV
env7
25 NHCNTSVITESCD
NTSVIT
NHCNTSVIQECCD
NTSVID
TSCNTSVITQACP
30 NTSVIT
INCNTSVITQACP
NTSVIT
INCNTSAITQACP
NTSAIT

env8

YCAPPGYALLRC-NDT

YCAPAGFAILKCNNKT

YCAPAGFAILKCNDKK

5 YCAPAGFAILKCRDKK

env9

NKRPRQAWCWFKG-KWKD

NERPKQAWCRFGG-NWKE

N--MRQAHCNISRAKWNA

10 D--IRRAYCTINETEWDK

I--IGQAHCNISRAQWSK

env10

KGSDPEVAYMWTNCRGEFLYCNMTWFLN

NCRGEFLYCN

15 GG-DPEVTFMWTNCRGEFLYCKMNWFLN

NCRGEFLYCK

-GGDPEIVTHSFNCGGEFFYCNSTQLFN

NCGGEFFYCN

-GGDPEITTHSFNCRGGEFFYCNTSKLFN

NCRGEFFYCN

20 -GGDPEITTHSFNCGGEFFYCNTSGLFN

NCGGEFFYCN

env11

RNYAPCHIKQIINTWHKVGRNVY

CHIKQII

25 RNYVPCHIRQIINTWHKVGVKNVY

CHIRQII

TITLPCRIKQFINMWQEVGKAMY

CRIKQFI

30 SITLPCRIKQIINMWQKTCKAMY

CRIKQII

NITLQCRICKQIIKVMAGR-KAIY

CRIKQII

qaq1

35 DCKLVLKGLGTNPTELLEMLTA

Les peptides selon l'invention peuvent encore avantageusement être préparés par les techniques classiques, dans le domaine de la synthèse des peptides. Cette synthèse peut être réalisée en solution homogène ou en phase solide.

Par exemple, on aura recours à la technique de synthèse en solution homogène décrit par Houbenweyl dans l'ouvrage intitulé "Méthode der Organischen Chemie" (Méthode de la Chimie Organique) édité par E. Wunsch, vol. 15-I et II., THIEME, Stuttgart 1974.

Cette méthode de synthèse consiste à condenser successivement deux-à-deux les aminoacyles successifs dans l'ordre requis, ou à condenser des aminoacyles et des fragments préalablement formés et contenant déjà plusieurs aminoacyles dans l'ordre approprié, ou encore plusieurs fragments préalablement ainsi préparés, étant entendu que l'on aura eu soin de protéger au préalable toutes les fonctions réactives portées par ces aminoacyles ou fragments, à l'exception des fonctions amines de l'un et carboxyles de l'autre ou vice-versa, qui doivent normalement intervenir dans la formation des liaisons peptidiques, notamment après activation de la fonction carboxyle, selon les méthodes bien connues dans la synthèse des peptides. En variante, on pourra avoir recours à des réactions de couplage mettant en jeu des réactifs de couplage classique, du type carbodiimide, tels que par exemple la 1-éthyl-3-(3-diméthyl-amino-propyl)-carbodiimide. Lorsque l'aminoacyle mis en œuvre possède une fonction acide supplémentaire (notamment dans le cas de l'acide glutamique), ces fonctions seront par exemple protégées, par des groupes t-bustylester.

Dans le cas de la synthèse progressive, acide aminé par acide aminé, la synthèse débute de préférence par la condensation de l'amino-acide C-terminal avec l'aminoacide qui correspond à l'aminoacyle voisin dans

la séquence désirée et ainsi de suite, de proche en proche, jusqu'à l'acide aminé N-terminal. Selon une autre technique préférée de l'invention, on a recours à celle décrite par R.D. MERRIFIELD dans l'article intitulé "Solid phase peptide synthesis" (J. Am. Soc., 5 45, 2149-2154).

Pour fabriquer une chaîne peptidique selon le procédé de MERRIFIELD, on a recours à une résine polymère très poreuse, sur laquelle on fixe le premier acide aminé C-terminal de la chaîne. Cet acide aminé est fixé 10 sur la résine par l'intermédiaire de son groupe carboxylique et sa fonction amine est protégée, par exemple par le groupe t-butyloxycarbonyle.

Lorsque le premier acide aminé C-terminal est 15 ainsi fixé sur la résine, on enlève le groupe protecteur de la fonction amine en lavant la résine avec un acide.

Dans le cas où le groupe protecteur de la fonction amine est le groupe t-butyloxycarbonyle, il peut être éliminé par traitement de la résine à l'aide 20 d'acide trifluoroacétique.

On couple ensuite le deuxième acide aminé qui fournit le second amino-acyle de la séquence recherché, à partir du résidu amino-acyle C-terminal sur la fonction amine déprotégée du premier acide aminé C-terminal fixé sur la chaîne. De préférence, la fonction carboxyle 25 de ce deuxième acide aminé est activée, par exemple par la dicyclohexylcarbodiimide, et la fonction amine est protégée, par exemple par le t-butyloxycarbonyle.

On obtient ainsi la première partie de la 30 chaîne peptidique recherchée, qui comporte deux acide aminés, et dont la fonction amine terminale est protégée. Comme précédemment, on déprotège la fonction amine et on peut ensuite procéder à la fixation du troisième aminoacyle, dans les conditions analogues à celles de l'addition du deuxième acide aminé C-terminal.

On fixe ainsi, les uns après les autres, les acides aminés qui vont constituer la chaîne peptidique sur le groupe amine chaque fois déprotégé au préalable de la portion de la chaîne peptidique déjà formée, et qui est rattachée à la résine.

5

Lorsque la totalité de la chaîne peptidique désirée est formée, on élimine les groupes protecteurs des différents acides aminés constituant la chaîne peptidique et on détache le peptide de la résine par exemple à l'aide d'acide fluorydrique.

10

L'invention concerne également les oligomères hydrosolubles des peptides monomères sus-indiqués. L'oligomérisation peut provoquer un accroissement de l'immunogénicité des peptides monomères selon l'invention. Sans qu'une telle indication chiffrée puisse être considérée comme limitative, on mentionnera néanmoins que ces oligomères peuvent, par exemple, contenir de 2 à 15 10 unités monomères.

20

Les unités monomères entrant dans cet oligomère sont soit toutes constituées par le polypeptide de séquence 1 ou par le polypeptide de séquence 2, soit par l'un et l'autre de ces polypeptides.

25

On peut avoir recours, pour réaliser l'oligomérisation, à toute technique de polymérisation couramment utilisée dans le domaine des peptides, cette polymérisation étant conduite jusqu'à l'obtention d'un oligomère ou polymère contenant le nombre de motifs monomères requis pour l'acquisition de l'immunogénicité désirée.

30

Une méthode d'oligomérisation ou de polymérisation du monomère consiste dans la réaction de celui-ci avec un agent de réticulation tel que le glutaraldéhyde.

35

On peut également avoir recours à d'autres méthodes d'oligomérisation ou de couplage, par exemple à

celle mettant en jeu des couplages successifs d'unités monomères, par l'intermédiaire de leurs fonctions terminales carboxyle et amine en présence d'agents de couplage homo- ou hétéro- bifonctionnels.

5 On peut également pour la production de molécules comportant un ou plusieurs motifs de 17 acides aminés tels que définis ci-dessus, avoir recours à des techniques du génie génétique mettant en oeuvre des micro-organismes transformés par un acide nucléique déterminé comprenant des séquences nucléotidiques appropriées correspondantes.
10

15 L'invention concerne également les acides nucléiques contenant une ou plusieurs séquences issues de la séquence de l'ADNc du virus HIV-2 ROD. Ces séquences repérées par la numérotation figurant sur la séquence précédemment décrite, codent pour certains peptides intéressants de l'invention.

Séquence codant pour env1 nucléotides 7850 à 7927

| | | | | | |
|----|---|---|--------------|---|-------------|
| | " | " | <u>env2</u> | " | 8030 à 8095 |
| 20 | " | " | <u>env3</u> | " | 7601 à 7636 |
| | " | " | <u>env4</u> | " | 6170 à 6247 |
| | " | " | <u>env5</u> | " | 6294 à 6349 |
| | " | " | <u>env6</u> | " | 6392 à 6445 |
| | " | " | <u>env7</u> | " | 6724 à 6763 |
| 25 | " | " | <u>env8</u> | " | 6794 à 6838 |
| | " | " | <u>env9</u> | " | 7112 à 7162 |
| | " | " | <u>env10</u> | " | 7253 à 7336 |
| | " | " | <u>env11</u> | " | 7358 à 7426 |
| | " | " | <u>gag1</u> | " | 1535 à 1597 |

30 L'invention concerne enfin les acides nucléiques correspondants du virus SIV, contenant une ou plusieurs séquences issues de l'ADNc du virus SIV-1. Ces séquences codant pour les peptides env1 à env11 et gag1 peuvent être repérés sur la figure 3 par comparaison avec les séquences correspondantes décrites pour HIV-2.
35

Il va de soi que l'invention concerne les acides nucléiques correspondant à des séquences placées en des régions analogues des ADNC dérivés de variants de HIV-2 ROD ou de SIV, ainsi que tous les acides nucléiques dont les modifications vis à vis des précédents résulteraient de la mise à profit de la dégénérescence du code génétique.

L'invention concerne encore les conjugués obtenus par couplage covalent des peptides selon l'invention (ou des susdits oligomères) à des molécules porteuses (naturelles ou synthétiques), physiologiquement acceptables et non toxiques, par l'intermédiaire de groupements réactifs complémentaires respectivement portés par la molécule porteuse et le peptide. Des exemples de groupements appropriés sont illustrés dans ce qui suit :

A titre d'exemple de molécules porteuses ou supports macromoléculaires entrant dans la constitution des conjugués selon l'invention, on mentionnera des protéines naturelles, telles que l'anatoxine tétanique, l'ovalbulmine, des sérum albumines, des hémocyanines, etc...

A titre de support macromoléculaires synthétiques, on mentionnera par exemple des polylysines ou des poly(D-L-alanine)-poly(L-lysine).

La littérature mentionne d'autres types de supports macromoléculaires susceptibles d'être utilisés, lesquels présentent en général un poids moléculaire supérieur à 20 000.

Pour synthétiser les conjugués selon l'invention, on peut avoir recours à des procédés connus en soi, tels que celui décrit par FRANTZ et ROBERTSON dans Infect. and Immunity, 33, 193-198 (1981), ou celui décrit dans Applied and Environmental Microbiology, (octobre 1981), vol. 42, n° 4, 611-614 par P.E. KAUFFMAN

29

en utilisant le peptide et la molécule porteuse appropriée.

Dans la pratique, on utilisera avantageusement comme agent de couplage les composés suivants, 5 cités à titre non limitatif : aldéhyde glutarique, chloroformiate d'éthyle, carbodiimides hydrosolubles [N-éthyl-N'(3-diméthylamino-propyl) carbodiimide, HCl], diisocyanates, bis-diazobenzidine, di- et trichloro-s-triazines, bromures de cyanogène, ainsi que les agents 10 de couplage mentionnés dans Scand. J. Immunol., (1978), vol. 8, p. 7-23 (AVRAMEAS, TERNYNCK, GUESDON).

On peut avoir recours à tout procédé de couplage faisant intervenir d'une part une ou plusieurs fonctions réactives du peptide et d'autre part, une ou 15 plusieurs fonctions réactives de molécules supports.

Avantageusement, il s'agit des fonctions carboxyle et amine, lesquelles peuvent donner lieu à une réaction de couplage en présence d'un agent de couplage du genre de ceux utilisés dans la synthèse des protéines, par 20 exemple, le 1-éthyl-3-(3-diméthylaminopropyl)-carbodiimide, le N-hydroxybenzotriazole, etc... On peut encore avoir recours à la glutaraldéhyde, notamment lorsqu'il s'agit de relier entre eux des groupes aminés respectivement portés par le peptide et la molécule 25 support.

Les peptides selon l'invention possèdent des propriétés antigéniques. Ils peuvent donc être utilisés dans des procédés de diagnostic pour la détection d'une infection par le virus HIV-2.

30 Comme on l'a déjà mentionné, des études ont permis de distinguer deux groupes de peptides pouvant être mis en oeuvre dans des procédés de détection d'anticorps contre le virus HIV-2 dans un fluide biologique humain, notamment un sérum ou un liquide céphalo-rachidien.

Un premier groupe (I) comprend les peptides gag₁. Ces peptides reconnaissent des anticorps anti-HIV-2 et sont donc capables de détecter une infection par HIV-2. Ils reconnaissent également dans une certaine mesure des anticorps anti-HIV-1.

5

Un second groupe (II) comprend des peptides qui correspondent plus particulièrement à ceux qui sont situés dans la partie transmembranaire et dans la fin de la partie externe de la protéine d'enveloppe. Ces peptides sont ceux précédemment désignés par env1, env2 et env3. Ils permettent la reconnaissance spécifique de la présence d'anticorps contre HIV-2 et permettent donc de discriminer chez une personne les infections passées ou présentes dues à un HIV, plus particulièrement entre celles qui ont été provoquées par un HIV-2 et celles qui 15 l'ont été par un HIV-1.

10

15

20

L'invention concerne également une composition contenant au moins l'un des susdits peptides ou au moins un oligomère de ce peptide, caractérisée en ce qu'elle a la capacité d'être reconnue par des sérums d'origine humaine contenant des anticorps contre le virus HIV-2.

25

L'invention concerne un procédé de diagnostic in vitro un ou des peptides selon l'invention pour la détection d'anticorps contre HIV-2 dans des fluides biologiques, en particulier dans des sérums humains.

D'une façon générale le procédé de diagnostic in vitro ci-dessus comprend les étapes suivantes :

30

- la mise en contact de ce liquide biologique avec lesdits peptides;
- la détection de la présence éventuelle d'un complexe peptidé-anticorps par des méthodes physiques ou chimiques, dans ledit liquide biologique.

35

Dans un mode de réalisation préféré de l'invention, la détection du complexe antigène-anticorps est réalisée grâce à des tests immunoenzymatiques (du type

ELISA), immunofluorescents (du type IFA), radioimmunologiques (du type RIA) ou des tests de radioimmunoprécipitation (du type RIPA).

Ainsi l'invention concerne également tout peptide selon 5 l'invention marqué à l'aide d'un marqueur adéquat du type enzymatique, fluorescent, radioactif, etc...

De telles méthodes comprennent par exemple les étapes suivantes :

- dépôt de quantités déterminées d'une 10 composition peptidique selon l'invention dans les puits d'une microplaquette de titration,

- introduction dans lesdits puits de dilutions croissantes du sérum devant être diagnostiqué,

- incubation de la microplaquette,

15 - rinçages répétés de la microplaquette,

- introduction dans les puits de la microplaquette d'anticorps marqués contre des immunoglobulines du sang, le marquage de ces anticorps ayant été réalisé à l'aide d'une enzyme sélectionnée parmi celles qui sont 20 capables d'hydrolyser un substrat en modifiant l'absorption des radiations de ce dernier, au moins à une longueur d'onde déterminée,

- détection, en comparaison avec un témoin de contrôle, de la quantité de substrat hydrolysé.

25 L'invention concerne également des coffrets ou kits pour le diagnostic in vitro de la présence d'anticorps contre les virus HIV-2 et, dans certains cas, HIV-1 dans un milieu biologique qui comprennent :

- une composition peptidique selon l'invention, 30

- les réactifs pour la constitution du milieu propice à la réalisation de la réaction immunologique,

- les réactifs permettant la détection du complexe antigènes-anticorps produit par la réaction immunologique. De tels réactifs peuvent également porter 35

un marqueur, ou être susceptibles d'être reconnus à leur tour par un réactif marqué. Plus particulièrement dans le cas où la composition polypeptidique sus-mentionnée n'est pas marquée.

5 - un tissu fluide biologique de référence dé-pourvu d'anticorps reconnus par la composition polypep-tidique sus-mentionnée,

L'invention concerne les anticorps eux-mêmes formés contre les peptides de l'invention.

10 Il va de soi que cette production n'est pas limitée aux anticorps polyclonaux.

Elle s'applique encore à tout anticorps mono-clonal produit par tout hybridome susceptible d'être formé, par des méthodes classiques, à partir des cel-lules spléniques d'un animal, notamment de souris ou de rat, immunisés contre l'un des peptides de l'invention, d'une part et des cellules d'une lignée de cellule myélome approprié d'autre part, et d'être sélectionné, par sa capacité à produire des anticorps monoclonaux reconnaissant le peptide initialement mis en oeuvre pour l'immunisation des animaux.

25 L'invention concerne également des composi-tions immunogènes pour la production de vaccins dont le principe actif est constitué par au moins un peptide selon l'invention, ou un oligomère de ce peptide, ou un peptide sous forme conjuguée avec une molécule porteuse, caractérisées en ce qu'elles induisent la production d'anticorps contre les susdits peptides en quantité suf-fisante pour aussi inhiber les protéines du rétrovirus 30 HIV-2, voire même le rétrovirus HIV-2 entrant en asso-ciation avec un véhicule pharmaceutiquement acceptable.

Les compositions immunogènes pour la produc-tion de vaccins comprennent de façon avantageuse plus particulièremenr au moins l'un des peptides précédemment désignés par env4, env5, env6, env7, env8, env9, env10,

env11 voir des mélanges de ceux-ci.

Parmi ces peptides aptes à constituer des principes actifs de vaccins certains sont particulièrement préférés car ils possèdent une structure de base en acides aminés correspondant à des régions des glycoprotéines d'enveloppe qui présentent un important degré de conservation, non seulement dans les HIV-2, et dans les SIV, mais également dans les HIV-1. Ces peptides particulièrement préférés sont les peptides désignés par 10 env4, certains peptides env5, env6 et env10.

Dans un mode de réalisation préféré de l'invention les peptides immunogènes (ou fragments de ces peptides) aptes à constituer des principes actifs de vaccins sont choisis parmi ceux dont les formules correspondent à des séquences qui, dans les glycoprotéines d'enveloppe de HIV-2, SIV et HIV-1 présentant une homologie en acides aminés supérieure à 50%, qui appartiennent à la partie externe de l'enveloppe du virus, qui sont dépourvus ou presque de délétions, et qui renferment des résidus de cystéine favorables à la stabilisation des liaisons et à la constitution de boucles d'ancre.

Les peptides suivants appartiennent à cette catégorie de peptides préférés.

25 env4

XVTV-YGVP-W--ATZ

env5

XL-NVTE-FZ

env6

30 XKPCVKL-PLC-Z

env7

XN-S-I-Z

env10

XNC-GEF-YC-Z

env11

XC-I-Q-IZ

Des compositions pharmaceutiques avantageuses sont constituées par des solutions, suspensions ou liposomes injectables contenant une dose efficace d'au moins un produit selon l'invention. De préférence, ces solutions, suspensions ou liposomes sont réalisés dans une phase aqueuse stérilisée isotonique, de préférence saline ou glucosée.

L'invention concerne plus particulièrement de telles suspensions, solutions ou forme liposome qui sont aptes à être administrées par injections intradermiques, intramusculaires ou sous-cutanées, ou encore par scarifications.

Elle concerne également des compositions pharmaceutiques administrables par d'autres voies, notamment par voie orale.

Les compositions pharmaceutiques selon l'invention, utilisables en tant que vaccins pour être efficaces dans la production d'anticorps contre le virus HIV-2, peuvent à titre d'exemple être administrées à des doses situées entre 10 et 500 µg/kg, de peptides selon l'invention, de préférence de 50 à 100 µg/kg.

Ces doses sont citées à titre d'exemple et ne possèdent en aucun cas un caractère limitatif.

Comme on l'a déjà indiqué plus haut les différents peptides qui ont été définis peuvent comprendre des modifications qui n'ont pas pour effet de modifier de façon fondamentale leurs propriétés immunologiques. Les peptides équivalents qui en résultent entrent dans le champ des revendications qui suivent. A titre d'exemples de peptides équivalents on mentionnera ceux dont les structures en correspondance avec des régions des ADNc d'autres variants de HIV-2 de SIV ou de HIV-1, lorsque ces régions ont été mises en alignement dans des

35

conditions semblables à celles qui ont été évoquées ci-dessus, à propos de HIV-2 ROD, SIV et HIV-1 BRU. A titre d'autres de ces peptides, on mentionnera ceux dont les structures sont en correspondance avec de telles régions dans les ADNc qui ont fait l'objet de dépôts à la CNCM, notamment sous les numéros I-502, I-642 (HIV-2 IRMO), I-643 (HIV-2 EHO) ainsi que, dans les cas appropriés, des variants de HIV-1 qui ont fait l'objet de dépôts à la CNCM sous les numéros I-232, I-240, I-241, I-550, I-551.

Les peptides selon l'invention peuvent encore être définis par les formules suivantes (dans lesquels X, Z et les tirets "--" ont les significations sus-indiquées) :

15

20

25

30

35

36

XRV-AIEKYL-DQA-LN-WGCAFROVCZ

XAIEKYL-DZ

X-LE-AQIQQEKNMYELQKLNSWZ

XQIQQEKNZ

5

XELGDXKLVEITPIG-APT--KR-----Z

XYKLVEITPIG-APT--KRZ

10 X----VTV-YGVP-W--AT--LFCA-Z

XVTV-YGVP-W--ATZ

X----E--L-NVTE-F--W-NZ

XL-NVTE-FZ

15

XL---S-KPCVKL-PLC----Z

XKPCVKL-PLC-Z

XS-KPCVKL-PLC-Z

20 X---N-S-I---C-Z

XN-S-I-Z

XYC-P-G-A-L-C-N-TZ

25 X-----A-C-----W--Z

NKRPRQAWCWFKG-KWKD

X-G-DPE-----NC-GEF-YC-----NZ

30

X-----C-I-Q-I-----G---YZ

35

• L'invention concerne également outre les peptides de SIV déjà décrits, les protéines codées par l'ADNc du virus SIV. Elle concerne également les protéines de tout virus immunologiquement étroitement apparenté à SIV-1mac, en particulier tout virus dont les protéines et les glycoprotéines d'enveloppe croisent immunologiquement et dont les ADNc présentent un pourcentage d'homologie d'au moins 95% et de préférence d'au moins 98%.

En particulier l'invention concerne :

- 10 1/ les protéines et glycoprotéines de l'enveloppe codées par le gène env et représentées à la figure 3,
- 2/ la protéine GAG représentée à la figure 4,
- 3/ la protéine POL représentée à la figure 5,
- 15 4/ la protéine Q représentée à la figure 6,
- 5/ la protéine R représentée à la figure 7,
- 6/ la protéine X représentée à la figure 8,
- 7/ la protéine F représentée à la figure 9,
- 8/ la protéine TAT représentée à la figure 10,

Les acides aminés des protéines précitées de SIV, ont été représentées en alignement avec les séquences d'acides aminés des protéines correspondantes du virus HIV-2 ; les points verticaux figurant entre les deux séquences correspondent aux acides aminés communs entre les protéines des deux virus.

25 Les séquences d'ADNc codant pour les protéines précitées apparaissent sur la figure 1B. L'invention concerne, outre les séquences nucléiques précitées toute séquence nucléique modifiée, qui code également pour les protéines du rétrovirus SIV ou d'un variant.

30 Ces séquences d'ADNc repérées par la numérotation figurant sur les séquences décrites précédemment (figure 1B) sont les suivantes :

| | | |
|----|---|--|
| | -séquence codant pour <u>GAG</u> , nucléotides 551 à 2068 | |
| | - " " <u>POL</u> , " 1726 à 4893 | |
| | - " " Q, " 4826 à 5467 | |
| | - " " X, " 5298 à 5633 | |
| 5 | - " " R, " 5637 à 5939 | |
| | - " " F, " 8569 à 9354 | |
| | - " " TAT-1 " 5788 à 6084 | |
| | - " " ART-1 " 6014 à 6130 | |
| | - " " TAT-2 " 8296 à 8391 | |
| 10 | - " " ART-2 " 8294 à 8548 | |
| | - " <u>ENV</u> " 6090 à 8732 | |

L'invention concerne donc naturellement les protéines précédemment décrites, lorsqu'elles sont obtenues à partir du virus SIV ou lorsqu'elles sont préparées par une méthode de synthèse, notamment par l'une des méthodes déjà citées en rapport avec la synthèse des peptides de plus petite taille.

L'invention concerne également l'utilisation des protéines précédentes pour le diagnostic de la présence éventuelle d'anticorps dirigés contre les protéines de HIV-2, voire contre HIV-2 en entier, ou pour certaines d'entre elles l'utilisation aux fins de diagnostic d'une infection due à l'un des virus HIV. Ainsi le peptide GAG codé par le gène correspondant peut être utilisé pour repérer la présence éventuelle d'anticorps anti-HIV-1 ou anti-HIV-2. Les protéines ENV sont utilisées de préférence pour le diagnostic spécifique d'une infection due à HIV-2 ou un de ses variants, parfois pour le diagnostic d'une infection par HIV-2 ou HIV-1.

L'invention concerne donc également un procédé de diagnostic in vitro de détection d'anticorps contre HIV-2 et éventuellement contre HIV-1 dans des fluides biologiques et en particulier dans des sérum humains. De tels procédés applicables pour l'utilisation des protéines précédentes de SIV comme protéines de diagnostic,

ont déjà été décrits dans la présente invention.

L'invention concerne aussi des coffrets ou "kits" pour le diagnostic in vitro de la présence d'anticorps le virus HIV-2 et dans certains cas contre HIV-1 dans un milieu biologique. De tels kits mettant en oeuvre les peptides précédents ont également été décrits dans la présente invention.

L'invention concerne également des compositions immunogènes pour la production de vaccins, dont le principe actif est constitué de façon avantageuse par au moins la partie de la protéine ENV du virus SIV, cette protéine pouvant être sous forme conjuguée avec une molécule porteuse. Ces compositions immunogènes induisent la production d'anticorps contre le susdit peptide en quantité suffisante pour inhiber les protéines du rétrovirus HIV-2, voire le rétrovirus HIV-2 lui-même.

Toutefois l'utilisation aux fins de diagnostic des protéines de SIV n'est en rien limitée à celle des seuls protéines ENV ou GAG. D'autres protéines parmi celles décrites peuvent être envisagées, pour préparer des compositions de diagnostic voire de vaccin.

25

30

35

REVENDICATIONS

- 1/ Peptide ayant des propriétés immunologiques en commun avec celles de l'ossature peptidique de la glycoprotéine d'enveloppe des virus de la classe HIV-2, caractérisé en ce qu'il a également une structure peptidique en commun avec l'ossature peptidique de la glycoprotéine de SIV.1.
- 2/ Peptide ayant des propriétés immunologiques en commun avec celles de l'ossature peptidique de la glycoprotéine d'enveloppe des virus de la classe HIV-2, ces peptides contenant un nombre de résidus d'acides aminés n'excédant pas 40, caractérisé en ce qu'il a également une structure peptidique en commun avec l'ossature peptidique de la glycoprotéine de SIV.1.
- 15 3/ Peptide selon la revendication 2 caractérisé par l'une des formules :
- XRV-AIEKYL-DQA-LN-WGCAFRQVCZ
- XAIEKYL-DZ
- dans laquelle X et Z sont des groupements OH ou NH₂ ou,
- 20 dans la mesure où les propriétés immunologiques du peptide dépourvu de ces groupes ne s'en trouvent pas essentiellement modifiées, des groupes comportant de 1 à 5 résidus d'acides aminés, et chacun des tirets correspond à un résidu aminoacyle choisi parmi ceux qui permettent de conserver au peptide sus-défini les propriétés immunologiques de l'une des séquences suivantes :
- 25 RVTAIEKYLQDQARLNNSWGCAFRQVC
- AIEKYLQDQ
- RVSAIEKYLKDQAQQLNAWGCAFRQVC
- 30 AIEKYLKDQ
- 4/ Peptide selon la revendication 2 caractérisé par l'une des formules :
- X-LE-AQIQQEKNMYELQKLNSWZ
- XQIQQEKNZ
- 35 dans laquelle X et Z sont des groupements OH ou NH₂ ou,

dans la mesure où les propriétés immunologiques du peptide dépourvu de ces groupes ne s'en trouvent pas essentiellement modifiées, des groupes comportant de 1 à 5 résidus d'acides aminés, et chacun des tirets correspond à un résidu aminoacyle choisi parmi ceux qui permettent de conserver au peptide sus-défini les propriétés immunologiques de l'une des séquences suivantes :

5 SLEQAQIQQEKNMYELQKLNSW

QIQQEKN

10 LLEEAQIQQEKNMYELQKLNSW

5/ Peptide selon la revendication 2 caractérisé par l'une des formules :

XELGDYKLVEITPIG-APT--KR-----Z

XYKLVEITPIG-APT--KRZ

15 dans laquelle X et Z sont des groupements OH ou NH₂ ou, dans la mesure où les propriétés immunologiques du peptide dépourvu de ces groupes ne s'en trouvent pas essentiellement modifiées, des groupes comportant de 1 à 5 résidus d'acides aminés, et chacun des tirets correspond à un résidu aminoacyle choisi parmi ceux qui permettent de conserver au peptide sus-défini les propriétés immunologiques de l'une des séquences suivantes :

20 ELGDYKLVEITPIGFAPTKEKRYSSAH

YKLVEITPIGFAPTKEK

25 ELGDYKLVEITPIGLAPTNVKRYTTG-
YKLVEITPIGLAPTNVK

6/ Peptide selon la revendication 2 caractérisé par l'une des formules :

30 X----VTV-YGVP-W--AT--LFCA-Z

XVTY-YGVP-W--ATZ

35 dans laquelle X et Z sont des groupements OH ou NH₂ ou, dans la mesure où les propriétés immunologiques du peptide dépourvu de ces groupes ne s'en trouvent pas essentiellement modifiées, des groupes comportant de 1 à 5 résidus d'acides aminés, et chacun des tirets correspond

42

à un résidu aminoacyle choisi parmi ceux qui permettent de conserver au peptide sus-défini les propriétés immunologiques de l'une des séquences suivantes :

CTQYVTVFYGVPTWKNATIPLFCAT

5 VTVFYGVPTWKNAT

CIQYVTVFYGVPAWRNATIPLFCAT

VTVFYGVPAWRNAT

EKLWVTVYYGVPVWKEATTLFCAS

VTVYYYGVPVWKEAT

10 7/ Peptide selon la revendication 6 caractérisé par l'une des formules :

CTQYVTVFYGVPTWKNATIPLFCAT

VTVFYGVPTWKNAT

CIQYVTVFYGVPAWRNATIPLFCAT

VTVFYGVPAWRNAT

EKLWVTVYYGVPVWKEATTLFCAS

VTVYYYGVPVWKEAT

EDLWVTVYYGVPVWKEATTLFCAS

VTVYYYGVPVWKEAT

20 DNLWVTVYYGVPVWKEATTLFCAS

VTVYYYGVPVWKEAT

8/ Peptide selon la revendication 2 caractérisé par l'une des formules :

X---E--L-NVTE-F--W-NZ

25 XL-NVTE-FZ

dans laquelle X et Z sont des groupements OH ou NH₂ ou, dans la mesure où les propriétés immunologiques du peptide dépourvu de ces groupes ne s'en trouvent pas essentiellement modifiées, des groupes comportant de 1 à 5 résidus d'acides aminés, et chacun des tirets correspond à un résidu aminoacyle choisi parmi ceux qui permettent de conserver au peptide sus-défini les propriétés immunologiques de l'une des séquences suivantes :

DDYQEITL-NVTEAFDAWNN

35 L-NVTE

DDYSELAL-NVTESFDAWEN

PNPQEVLVNVTENFNMWKN

LVNVTE

9/ Peptide selon la revendication 8 caractérisé
5 par l'une des formules :

DDYQEITL-NVTEAFDAWNN

L-NVTEAF

DDYSELAL-NVTESFDAWEN

L-NVTESF

10 PNPQEVLVNVTENFNMWKN

LVNVTENF

PNPQEIELENVTEGFNMWKN

LENVTEGF

PNPQEIALENVTENFNMWKN

LENVTENF

15 10/ Peptide selon la revendication 2 caractérisé

par l'une des formules :

XL---S-KPCVKL-PLC---Z

XKPCVKLTPLCVZ

XS-KPCVKLTPLCVZ

20 dans laquelle X et Z sont des groupements OH ou NH₂ ou,
dans la mesure où les propriétés immunologiques du pep-
tide dépourvu de ces groupes ne s'en trouvent pas essen-
tiellement modifiées, des groupes comportant de 1 à 5
25 résidus d'acides aminés, et chacun des tirets correspond
à un résidu aminoacyle choisi parmi ceux qui permettent
de conserver au peptide sus-défini les propriétés immu-
nologiques de l'une des séquences suivantes :

LFETSIKPCVKLTPLCVAMK

30 LFETSIKPCVKLSPLCITMR

LWDQSLKPCVKLTPLCVSLK

KPCVKLTPLCV

KPCVKLSPLCI

SLKPCVKLTPLCV

44

11/ Peptide selon la revendication 10 caractérisé par l'une des structures suivantes :

LFETSIKPCVKLTPLCVAMK

LFETSIKPCVKLSPLCITMR

5 LWDQSLKPCVKLTPLCVSLK

LWDQSLKPCVKLTPLCVTLN

PCVKLTPLCV

KPCVKLSPLCI

12/ Peptide selon la revendication 2 caractérisé en ce qu'il contient la structure de base :

10 X---N-S-I---C-Z

XN-S-I-Z

15 dans laquelle X et Z sont des groupements OH ou NH₂ ou, dans la mesure où les propriétés immunologiques du peptide dépourvu de ces groupes ne s'en trouvent pas essentiellement modifiées, des groupes comportant de 1 à 5 résidus d'acides aminés, et chacun des tirets correspond à un résidu aminoacyle choisi parmi ceux qui permettent de conserver au peptide sus-défini les propriétés immunologiques de l'une des séquences suivantes :

20 NHCNTSVITESCD

NTSVIT

NHCNTSVIQECCD

NTSVIQ

25 TSCNTSVITQACP

NTSVIT

13/ Peptide selon la revendication 12 caractérisé par l'une des formules suivantes :

NHCNTSVITESCD

30 NTSVIT

NHCNTSVIQECCD

NTSVIQ

TSCNTSVITQACP

NTSVIT

35 INCNTSVITQACP

NTSVIT

INCNTSAITQACP

NTSAIT

- 14/ Peptide selon la revendication 2 caractérisé
5 par l'une des formules suivantes :

XYC-P-G-A-L-C-N-TZ

dans laquelle X et Z sont des groupements OH ou NH₂ ou,
dans la mesure où les propriétés immunologiques du
peptide dépourvu de ces groupes ne s'en trouvent pas es-
10 sentiellement modifiées, des groupes comportant de 1 à 5
résidus d'acides aminés, et chacun des tirets correspond
à un résidu aminoacyle choisi parmi ceux qui permettent
de conserver au peptide sus-défini les propriétés immu-
nologiques de l'une des séquences suivantes :

15 YCAPPGYALLRC-NDT

YCAPAGFAILKCNNKT

- 15/ . Peptide selon la revendication 14 caractérisé
par l'une des formules :

YCAPPGYALLRC-NDT

20 YCAPAGFAILKCNNKT

YCAPAGFAILKCNDKK

YCAPAGFAILKCRDKK

- 16/ Peptide selon la revendication 2 caractérisé
par la formule :

25 X-----A-C-----W--Z

dans laquelle X et Z sont des groupements OH ou NH₂ ou,
dans la mesure où les propriétés immunologiques du
peptide dépourvu de ces groupes ne s'en trouvent pas es-
30 sentiellement modifiées, des groupes comportant de 1 à 5
résidus d'acides aminés, et chacun des tirets correspond
à un résidu aminoacyle choisi parmi ceux qui permettent
de conserver au peptide sus-défini les propriétés immu-
nologiques de l'une des séquences suivantes :

NKRPRQAWCWFKG-KWKD

35 NERPKQAWCRFGG-NWKE

N--MRQAHCNISRAKWNA

17/ Peptide selon la revendication 16 caractérisé par la formule suivante :

NKRPRQAWCWFKG-KWKD

5 NERPKQAWCRFGG-KWKE

N--MRQAHCNISRAKWNA

D--IRRAYCTINETEWDK

I--IGQAHCNISRAQWSK

18/ Peptide selon la revendication 2 caractérisé 10 par la formule suivantes :

X-G-DPE-----NC-GEF-YC-----N2

XNC-GEF-YC-Z

15 dans laquelle X et Z sont des groupements OH ou NH₂ ou, dans la mesure où les propriétés immunologiques du peptide dépourvu de ces groupes ne s'en trouvent pas essentiellement modifiées, des groupes comportant de 1 à 5 résidus d'acides aminés, et chacun des tirets correspond à un résidu aminoacyle choisi parmi ceux qui permettent de conserver au peptide sus-défini les propriétés immunologiques de l'une des séquences suivantes :

20 KGSDPEVAYMWTNCRGEFLYCNMTWFLN

NCRGEFLYCN

GG-DPEVTFMWTNCRGEFLYCKMNWFLN

NCRGEFLYCK

25 -GGDPEIVTHSFNCGGEFFYCNSTQLFN

NCGGEFFYCN

19/ Peptide selon la revendication 18 caractérisé 30 par l'une des structures suivantes :

KGSDPEVAYMWTNCRGEFLYCNMTWFLN

NCRGEFLYCN

GG-DPEVTFMWTNCRGEFLYCKMNWFLN

NCRGEFLYCK

-GGDPEIVTHSFNCGGEFFYCNSTQLFN

NCGGEFFYCN

35 -GGDPEITTHSFNCRGEFFYCNTSKLFN

NCRGEFFYCN

-GGDPEITTHSFNCGGEFFYCNTSGLFN

NCGGEFFYCN

- 20/ Peptide selon la revendication 2 caractérisé
5 par l'une des formules suivantes :

X-----C-I-Q-I-----G---YZ

XC-I-Q-IZ

10 dans laquelle X et Z sont des groupements OH ou NH₂ ou,
dans la mesure où les propriétés immunologiques du
peptide dépourvu de ces groupes ne s'en trouvent pas es-
sentiellement modifiées, des groupes comportant de 1 à 5
résidus d'acides aminés, et chacun des tirets correspond
à un résidu aminoacyle choisi parmi ceux qui permettent
de conserver au peptide sus-défini les propriétés immu-
nologiques de l'une des séquences suivantes :

15 RNYAPCHIKQIINTWHKVGRNVY

CHIKQII

RNYVPCHIRQIINTWHKVGVKNVY

CHIRQII

20 TITLPCRIKQFINMWQEVGKAMY

CRIKQFI

- 21/ Peptide selon la revendication 20 caractérisé
par l'une des structures suivantes :

RNYAPCHIKQIINTWHKVGRNVY

CHIKQII

25 RNYVPCHIRQIINTWHKVGVKNVY

CHIRQII

TITLPCRIKQFINMWQEVGKAMY

CRIKQFI

30 SITLPCRIKQIINMWQKTCKAMY

CRIKQII

NITLQCRIKQIIKMVAGR-KAIY

CRIKQII

- 35 22/ Peptide antigénique gag1, caractérisé par
l'une des structures de base :

XDCKLVLKGLGMNPTLEEMLTAZ

XDCKLVLKGLGTNPTLEEMLTAZ

dans lesquelles X et Z sont des groupements OH ou NH₂
ou, dans la mesure où les propriétés immunologiques du
5 peptide dépourvu de ces groupes ne s'en trouvent pas es-
sentiellement modifiées, des groupes comportant de 1 à 5
résidus d'acides aminés, et dans lesquelles chacun des
tirets correspond à un résidu aminoacyle choisi parmi
ceux qui permettent de conserver au peptide sus-défini
10 les propriétés immunologiques de l'une ou l'autre des
séquences suivantes :

DCKLVLKGLGMNPTLEEMLTA

DCKLVLKGLGTNPTLEEMLTA

23/ Séquence nucléotidique caractérisée en ce
15. qu'elle renferme tout ou partie de la séquence d'acides
nucléiques définie à la figure 1B.

24/ Séquence nucléotidique caractérisée en ce
qu'elle renferme tout ou partie de la séquence d'acides
nucléiques définie à la figure 1C.

20 25/ Séquence nucléotidique selon la revendication
23, caractérisée en ce qu'elle comprend les séquences
nucléotidiqes :

GAG s'étendant entre les nucléotides 550 à 2068

| | | | | | |
|----|-------|---|---|---|-----------------------|
| | POL | " | " | " | 1726 à 4893 |
| 25 | Q | " | " | " | 4826 à 5467 |
| | X | " | " | " | 5298 à 5633 |
| | R | " | " | " | 5637 à 5939 |
| | F | " | " | " | 8569 à 9354 |
| | TAT-1 | " | " | " | 5788 à 6084 |
| 30 | ART-1 | " | " | " | 6014 à 6130 |
| | TAT-2 | " | " | " | 8296 à 8391 |
| | ART-2 | " | " | " | 8294 à 8548 |
| | LTR | " | " | " | 8950 à 9468 <u>et</u> |
| | | | | | 1 à 316 |
| 35 | ENV | " | " | " | 6090 à 8732 |

26/ Peptide ayant une structure peptidique en commun avec l'ossature peptidique de SIV-1, caractérisé en ce qu'il comprend tout ou partie des séquences d'acides aminés parmi les séquences suivantes :

5 ENV représentée à la figure 3

| | | | |
|------------|---|---|----|
| <u>GAG</u> | " | " | 4 |
| <u>POL</u> | " | " | 5 |
| Q | " | " | 6 |
| R | " | " | 7 |
| 10 X | " | " | 8 |
| F | " | " | 9 |
| TAT | " | " | 10 |
| ART | " | " | 11 |

15 27/ Acide nucléique recombinant caractérisé en ce qu'il comprend la totalité ou une partie d'un ADNc selon 16 l'une quelconque des revendications 23 à 25, inséré dans un acide nucléique provenant d'un vecteur.

28/ Acide nucléique recombinant selon la revendication 27, caractérisé en ce qu'il est marqué.

20 29/ Composition antigénique contenant le peptide gag selon la revendication 26 ou 27, au moins un peptide gag1 selon la revendication 22 ou au moins un oligomère de ce peptide, caractérisée en ce qu'elle a la capacité d'être reconnue par des fluides biologiques d'origine humaine, notamment des sérums contenant des anticorps anti-HIV-2 et dans une certaine mesure des anticorps anti-HIV-1.

30 30/ Composition antigénique contenant le peptide env selon la revendication 26 ou au moins un peptide selon les revendications 3, 4 et 5 ou au moins un oligomère de ce peptide, caractérisée en ce qu'elle reconnaissent spécifiquement la présence d'anticorps contre HIV-2.

35 31/ Composition immunogène contenant tout ou partie du peptide env selon la revendication 26 ou au moins

un peptide ou au moins un oligomère de ce peptide ou ce peptide sous forme conjuguée avec une molécule porteuse, selon les revendications 6 à 21, en association avec un véhicule pharmaceutique acceptable pour la production de vaccins, caractérisée en ce qu'elle induit la production d'anticorps contre les susdits peptides en quantité suffisante pour inhiber efficacement les protéines du rétrovirus HIV-2, voire même le rétrovirus HIV-2 entier.

32/ Composition immunogène selon la revendication 10 caractérisée en ce qu'elle contient les peptides dont les formules correspondent à des séquences qui, dans les glycoprotéines d'enveloppe de HIV-2, SIV-1 et HIV-1 présentent une homologie en acides aminés supérieure à 50%.

15 33/ Composition immunogène selon l'une des revendications 31 ou 32, caractérisée en ce qu'elle contient au moins un peptide ou au moins un oligomère de ce peptide ou ce peptide sous forme conjuguée avec une molécule porteuse choisi parmi env4, env5, env6 et env10.

20 34/ Procédé de diagnostic in vitro de l'infection par HIV-2 dans un liquide biologique comprenant :

- la mise en contact de ce liquide biologique avec au moins un peptide selon l'une des revendications 1, 2, 3, 4, 5, 22 ou un conjugué de ces peptides avec une molécule porteuse ou des peptides gag ou env selon la revendication 26.

- la détection de la présence éventuelle d'un complexe antigène-anticorps par des méthodes physiques ou chimiques, dans ledit liquide biologique.

30 35/ Procédé de diagnostic in vitro de l'infection par HIV-2 dans un liquide biologique selon la revendication 34, caractérisé en ce que la détection du complexe antigène-anticorps éventuellement formé est réalisée grâce à des tests immunoenzymatiques (du type

51

ELISA) immunofluorescents (du type IFA) radioimmunologiques (du type RIA) ou des tests de radioimmunoprécipitation (du type RIPA).

36/ Kit pour le diagnostic in vitro de l'infection par HIV-2 dans un liquide biologique caractérisé en ce qu'il comprend :

- une composition peptidique contenant un peptide selon l'une des revendications 1 à 5, 22, ou un mélange de ces peptides, ou un conjugué de ces peptides avec une molécule porteuse, ou les peptides gag ou env selon la revendication 26,
- un réactif pour la constitution du milieu propice à la réalisation d'une réaction immunologique,
- un ou plusieurs réactifs éventuellement marqué pour la détection du complexe antigène-anticorps formé par la réaction immunologique,
- un liquide biologique de référence dépourvu d'anticorps reconnus par la susdite composition peptidique.

20

25

30

35

FIG. 1.A

HIV2.ROD

R →
 GTCGCTCTCGGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCACTAGCAGG
 TAGAGCCTGGGTGTTCCCTGCTAGACTCTACCCAGCACTGGCCGGTGTGGCAGACGG
 100 CCCACGCTTGCTTGCCTAAACCTCTTAAATAAGCTGCCAGTTAGAAGCAAGTTAAGT
 R US
 GTCTGCTCCATCTCTCTAGTCGCCGCTGGTCATTGGTGTTCACCTGAGTAACAAGA
 200 CCTGGTCTGTTAGGACCCTCTGCTTGGAAACCGAGGCAGGAAAATCCCTAGCAGG
 300 TTGGCCCTGAACAGGGACTTGAGAAGACTGAGAAGTCTTGGAACACGGCTGAGTGAAG
 GCAGTAAGGGCGGCAGGAACAAACCACGACGGAGTGCTCCTAGAAAGGCCGGGGAGG
 400 CACCAAAGGCAGCGTGTGGAGCGGGAGGAGAACAGGCCCCGGTGAAGGTAAGTACCTA
 CACCAAAAATGTAGCCGAAAGGGCTTGCTATCCTAACCTTACACAGCTAGAAGATTGTG
 500 MetGlyAlaArgAsnSerValLeuArgGlyLysLysAlaAspGluLeuGluArgIle
 GGAGATGGCGCGAGAAACTCCGCTTGAAGGGAAAAAGCAGATGAATTAGAAAGAAT
 600 ArgLeuArgProGlyGlyLysLysTyrArgLeuLysHisIleValTrpAlaAlaAsn
 CAGGTTACGGCCCGGGAAAGAAAAAGTACAGGCTAAACATATTGTGTGCGCAGCGAA
 LysLeuAspArgPheGlyLeuAlaGluSerLeuLeuGluSerLysGluGlyCysGlnLys
 TAAATTGGACAGATTGGATTAGCAGAGAGCCTGTTGAGTCAGGCTAAAGAGGGTTGTCAAA
 700 IleLeuThrValLeuAspProMetValProThrGlySerGluAsnLeuLysSerLeuPhe
 ATTCTTACAGTTTAGCTTACAGCTTACAGGTCACGGTACAGGTTCAAGAAAATTAAAGCTTTT
 AsnThrValCysValIleTrpCysIleHisAlaGluGluLysValLysAspThrGluGly
 TAATACTGTCGCGTCAATTGGTGCATACACGCAGAAGAGAAAGTGAAGATACTGAAAG
 800 AlaLysGlnIleValArgArgHisLeuValAlaGluThrGlyThrAlaGluLysMetPro
 AGCAAAACAAATAGTGGGAGACATCTAGTGGCAGAACAGGAACAGGAACAGGAAATGCC

FIG. 1A

2/35

SerThrSerArgProThrAlaProSerSerGluLysGlyGlyAsnTyrProValGlnHis
 AAGCACAAGTAGACCAACAGCACCATCTAGCGAGAAGGGAGGAAATTACCCACTGCAACA
 ValGlyGlyAsnTyrThrHisIleProLeuSerProArgThrLeuAsnAlaTrpValLys
 TGTAGGGGGCAACTACACCCATATAACCGCTGAGTCCCCGAACCCATAATGCCCTGGGTAAA
 1000
 LeuValGluGluLysLysPheGlyAlaGluValValProGlyPheGlnAlaLeuSerGlu
 ATTAGTAGAGGAAAAAAAGTTGGGGCAGAAGTAGTGGCAGGATTCAGGCCTCTCAGA
 GlyCysThrProTyrAspIleAsnGlnMetLeuAsnCysValGlyAspHisGlnAlaAla
 AGGCTGCACGCCCTATGATATCAACCAAATGCTTAATTGTGTGGGCACCATAAGCAGC
 1100
 MetGlnIleIleArgGluIleIleAsnGluGluAlaAlaGluTrpAspValGlnHisPro
 CATGCAGATAATCAGGGAGATTATCAATGAGGAAGCAGCAGAATGGGATGTGCAACATCC
 1200
 IleProGlyProLeuProAlaGlyGlnLeuArgGluProArgGlySerAspIleAlaGly
 AATACCAGGCCCCCTTACCGGGGGCAGCTTAGAGAGCCAAGGGATCTGACATAGCAGG
 ThrThrSerThrValGluGlnIleGlnTrpMetPheArgProGlnAsnProValPro
 GACAACAACACAGTAGAACAGATCCACTGGATGTTAGGCCACAAAATCCTGTACC
 1300
 ValGlyAsnIleTyrArgArgTrpIleGlnIleGlyLeuGlnLysCysValArgMetTyr
 AGTAGGAAACATCTATAGAACATGGATCCAGATAGGATTGAGAAGTGTGTCAGGATGTA
 AsnProThrAsnIleLeuAspIleLysGlnGlyProLysGluProPheGlnSerTyrVal
 CAACCCGACCAACATCCTAGACATAAAACAGGGACCAAGGAGCCGTTCCAAGCTATGT
 1400
 AspArgPheTyrLysSerLeuArgAlaGluGlnThrAspProAlaValLysAsnTrpMet
 AGATAGATTCTACAAAGCTTGAGGGCAGAACAAACAGATCCAGCAGTGAAGAATTGGAT
 1500
 ThrGlnThrLeuLeuValGlnAsnAlaAsnProAspCysLysLeuValLeuLysGlyLeu
 GACCCAAACACTGCTAGTACAAATGCCAACCCAGACTGTAAATTAGTGTCTAAAGGACT
 GlyMetAsnProThrLeuGluGluMetLeuThrAlaCysGlnGlyValGlyGlyProGly
 AGGGATGAACCTACCTAGAACAGATGCTGACCGCCCTGTCAGGGGTAGGTGGCCAGG
 1600
 GlnLysAlaArgLeuMetAlaGluAlaLeuLysGluValIleGlyProAlaProIlePro
 CCAGAAAGCTAGATTAATGCCAGAGGCCCTGAAAGAGGTCACTGGACCTGCCCTATCCC
 PheAlaAlaAlaGlnGlnArgLysAlaPheLysCysTrpAsnCysGlyLysGluGlyHis
 ATTCGCAGCAGCCCAGCAGAGAACAGCATTAAATGCTGGAACGTGGAAAGGAAGGGCA
 1700
 SerAlaArgGlnCysArgAlaProArgArgGlnGlyCysTrpLysCysGlyLysProGly
 CTGGCAAGACAATGCCGAGCACCTAGAACGGCAGGGCTGCTGGAAGTGTGGTAAGCCAGG
 1800
 ThrGlyArgPhePheArgThrGlyProLeuGly
 HisIleMetThrAsnCysProAspArgGlnAlaGlyPheLeuGlyLeuGlyProTrpGly
 ACACATCATGACAAACTGCCAGATAGACAGGCAGGTTTTAGGACTGGGCCCTGGGG
 LysGluAlaProGlnLeuProArgGlyProSerSerAlaGlyAlaAspThrAsnSerThr
 LysLysProArgAsnPheProValAlaGlnValProGlnGlyLeuThrProThrAlaPro
 AAAGAAGCCCCGCAACTTCCCCGTGGCCCAAGTCCCGCAGGGCTGACACCAACAGCACC
 1900
 ProSerGlySerSerSerGlySerThrGlyGluIleTyrAlaAlaArgGluLysThrGlu
 ProValAspProAlaValAspLeuLeuGluLysTyrMetGlnGlnGlyLysArgGlnArg
 CCCAGTGGATCCAGGAGTGGACTACTGGAGAAATATATGCAGCAAGGGAAAAGACAGAG
 ArgAlaGluArgGluThrIleGlnGlySerAspArgGlyLeuThrAlaProArgAlaGly
 GluGlnArgGluArgProTyrLysGluValThrGluAspLeuLeuHisLeuGluGlnGly
 AGAGCAGAGAGAGACCATACAAGGAAGTGACAGAGGACTTACTGCACCTCGAGCAGGG
 (fig. 1a-suite k)

3/35

GlyAspThrIleGlnGlyAlaThrAsnArgGlyLeuAlaAlaProGinPheSerLeuTrp
 GluThrProTyrArgGluProProThrGluAspLeuLeuHisLeuAsnSerLeuPheGly
 GGAGACACCATAACAGGGAGCCACCAACAGAGGACTTGCTGCACCTCAATTCTCTTTGG
 2100
 LysArgProValValThrAlaTyrIleGluGlyGlnProValGluValLeuLeuAspThr
 LysAspGln
 AAAAGACCACTAGTCACAGCATAACATTGAGGGTCAGCCAGTAGAAGTCTTGTAGACACA
 .
 GlyAlaAspAspSerIleValAlaGlyIleGluLeuGlyAsnAsnTyrSerProLysIle
 GGGGCTGACGACTCAATAGTAGCAGGAATAGAGTTAGGGAAACAATTATAGCCCCAAAATA
 2200
 ValGlyGlyIleGlyGlyPheIleAsnThrLysGluTyrLysAsnValGluIleGluVal
 GTAGGGGAATAGGGGATTCAATAAACCAAGGAATATAAAATGTAGAAATAGAAAGTT
 .
 LeuAsnLysLysValArgAlaThrIleMetThrGlyAspThrProIleAsnIlePheGly
 CTAAATAAAAAGGTACGGGCCACCATAATGACAGGCACACCCCAATCAACATTTGGC
 2300
 ArgAsnIleLeuThrAlaLeuGlyNetSerLeuAsnLeuProValAlaLysValGluPro
 AGAAATATTCTGACAGCCTTAGGCATGTCATTAAATCTACCAGTCGCCAAAGTAGGCCA
 2400
 IleLysIleMetLeuLysProGlyLysAspGlyProLysLeuArgGlnTrpProLeuThr
 ATAAAAATAATGCTAAAGCCAGGCAAAGATGGACCAAAACTGAGACAATGGCCCTTAACA
 .
 LysGluLysIleGluAlaLeuLysGluIleCysGluLysMetGluLysGluGlyGlnLeu
 AAAGAAAAAAATAGAACGACTAAAACAAATCTGTGAAAAAAATGGAAAAAGAAGGCCAGCTA
 2500
 GluGluAlaProProThrAsnProTyrAsnThrProThrPheAlaIleLysLysAsp
 GAGGAAGCACCTCCAACTAATCCTTATAATACCCCCACATTGCAATCAAGAAAAAGGAC
 .
 LysAsnLysTrpArgMetLeuIleAspPheArgGluLeuAsnLysValThrGlnAspPhe
 AAAAACAAATGGAGGATGCTAATAGATTCAGAGAACTAAACAAGGTAACCAAGATTTC
 2600
 ThrGluIleGlnLeuGlyIleProHisProAlaGlyLeuAlaLysLysArgArgIleThr
 ACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGGTTGCCAAGAAGAGAAGAAATTACT
 2700
 ValLeuAspValGlyAspAlaTyrPheSerIleProLeuHisGluAspPheArgProTyr
 GTACTAGATGTAGGGATGCTACTTCCATACCACTACATGAGGACTTTAGACCATA
 .
 ThrAlaPheThrLeuProSerValAsnAsnAlaGluProGlyLysArgTyrIleTyrLys
 ACTGCATTACTCTACCATCAGTGAACAATGCAGAACCCAGGAAAAAGATACTATATAAA
 2800
 ValLeuProGinGlyTrpLysGlySerProAlaIlePheGlnHisThrMetArgGlnVal
 GTCTTGCACAGGGATGCAAGGATCACCAAGCAATTTCACACACAATGAGACAGGTA
 .
 LeuGluProPheArgLysAlaAsnLysAspValIleIleGlnTyrMetAspAspIle
 TTAGAACCATTCAGAAACCAAGGATGTCATTATCATTGAGTACATGGATGATATC
 2900
 LeuIleAlaSerAspArgThrAspLeuGluHisAspArgValValLeuGinLeuLysGlu
 TTAATAGCTAGTGACAGGACAGATTAGAACATGAGGCTAGTCCCTGCAGCTCAAGGAA
 3000
 LeuLeuAsnGlyLeuGlyPheSerThrProAspGluLysPheGlnLysAspProProTyr
 CTTCTAAATGGCCTAGGATTTCTACCCAGATGAGAACGTTCCAAAAGACCTCCATAC
 .
 HisTrpMetGlyTyrGluLeuTrpProThrLysTrpLysLeuGlnLysIleGlnLeuPro
 CACTGGATGGGCTATGAACTATGCCAACTAAATGGAAGTTGCAGAAAATACAGTTGCC
 3100
 GlnLysGluIleTrpThrValAsnAspIleGlnLysLeuValGlyValLeuAsnTrpAla
 CAAAAAGAAATATGGACAGTCATGACATCCAGAAGCTAGTGGGTGTCCATAATTGGGCA

(fig.1A-suite 2)

4/35

AlaGlnIleTyrProGlyIleLysThrLysHisLeuCysArgLeuIleArgGlyLysMet
 GCACAACTCTACCCAGGGATAAAAGACCAAACACTATGTAGGTTAACAGAGAAAAATG
 3200
 ThrLeuThrGluGluValGlnTrpThrGluLeuAlaGluLeuGluAsnArg
 ACACTCACAGAAGAAGTACAGTGGACAGAATTAGCAGAAGCAGAGCTAGAAGAAAACAGA
 3300
 IleIleLeuSerGlnGluGlnGluGlyHisTyrTyrGlnGluGluLysGluLeuGluAla
 ATTATGCCAGGAACAAGAGGGACACTATTACCAAGAAGAAAAGAGCTACAGAACGA

 ThrValClnLysAspGluGluAsnGluTrpThrTyrLysIleHisGlnGluGluLysIle
 AGAGTCCAAAAGGATCAAGAGAATGAGTGGACATATAAACACCAGGAAGAAAAAATT

 LeuLysValGlyLysTyrAlaLysValLysAsnThrHisThrAsnGlyIleArgLeuLeu
 CTAAAAGTAGGAAAATATGCAAAGGTGAAAAACACCCATACCAATGGAATCAGATTGTTA

 AlaGlnValValGlnLysIleGlyLysGluAlaLeuValIleTrpGlyArgIleProLys
 GCACAGGTAGTCAGAAAATAGGAAAAGAAGCAGTACTAGTCATTGGGGACGAATACCAAAA
 3500
 PheHisLeuProValGlnArgGluIleTrpGluGlnTrpTrpAspAsnTyrTrpGlnVal
 TTTCACCTACCACTAGAGAGAGAAATCTGGGAGCAGTGGTGGGATAACTACTGGCAAGTG
 3600
 ThrTrpIleProAspTrpAspPheValSerThrProProLeuValArgLeuAlaPheAsn
 ACATGGATCCCAGACTGGGACTTCGTGCTACCCCACCACTGGTCAGGTTAGCGTTAAC

 LeuValGlyAspProIleProGlyAlaGluThrPheTyrThrAspGlySerCysAsnArg
 CTGGTAGGGGATCCTATACCAAGGTGCAGAGACCTCTACACAGATGGATCCTGCAATAGG
 3700
 GlnSerLysGluGlyLysAlaGlyTyrValThrAspArgGlyLysAspLysValLysLys
 CAATCAAAAGAAGGAAAAGCAGGATATGTAACAGATAAGGGAAAAGACAAGGTAAAGAAA

 LeuGluGlnThrThrAsnGlnAlaGluLeuGluAlaPheAlaMetAlaLeuThrAsp
 CTAGAGCAAATACCAATCAGCAAGCAGAACTAGAACGCTTGCATGGCACTAACAGAC
 3800
 SerGlyProLysValAsnIleIleValAspSerGlnTyrValMetGlyIleSerAlaSer
 TCGGGTCCAAAAGTTAATATTATAGTAGACTCACACTATGTAATGGGGATCAGTGCAGC
 3900
 GlnProThrGluSerGluSerLysIleValAsnGlnIleIleGluGluMetIleLysLys
 CAACCAACAGAGTCAGAAAGTAAATAGTGAACCAGATCATAGAAGAAAATGATAAAAAAG

 GluAlaIleTyrValAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlnGluVal
 GAAGCAATCTATGTTGCATGGGTCCCAGCCCCACAAAGGCATAGGGGGAAACCAGGAAGTA
 4000
 AspHisLeuValSerGlnGlyIleArgGlnValLeuPheLeuGluLysIleGluProAla
 GATCATTTAGTGAGTCAGGGTATCAGACAAGTGTGTTCCCTGGAAAAAATAGAGCCCCGT

 GlnGluGluHisGluLysTyrHisSerAsnValLysGluLeuSerHisLysPheGlyIle
 CAGGAAGAACATGAAAAATATCATAGCAATGTAAGAAACTGTCTCATAAATTGGAAATA
 4100
 ProAsnLeuValAlaArgGlnIleValAsnSerCysAlaGlnCysGlnGlnLysGlyGlu
 CCCAATTAGTGGCAAGGAAATAGTAAACTCATGTGCCCAATGTCAACAGAAAGGGAA
 4200
 AlaiIleHisGlyGlnValAsnAlaGluLeuGlyThrTrpGlnMetAspCysThrLysLeu
 GCTATACATGGGCAAGTAAATGCAGAACTAGGCACCTGGCAAATGGACTGCACACATTAA

 GluGlyLysIleIleIleValAlaValHisValAlaSerGlyPheIleGluAlaGluVal
 GAAGGAAAGATCATTATAGTAGCAGTACATGTTGCAAGTGGATTATAGAAGCAGAACG
 4300
 IleProGlnGluSerGlyArgGlnThrAlaLeuPheLeuLeuLysLeuAlaSerArgTrp
 ATCCCACAGGAATCAGGAAGACAAACAGCAGCACTGTTCTATTGAAACTGGCAAGTAGGTGG
 (fig.1A-suite 3)

5/35

ProIleThrHisLeuHisThrAspAsnGlyAlaAsnPheThrSerGlnGluValEysMet
 CCAATAACACACACTTGCATACAGATAATGGTGCCAACTTCACAGGAGGTGAAGATG
 4400 . .
 ValAlaTrpTrpIleGlyIleGluGlnSerPheGlyValProTyrAsnProGlnSerGln
 GTAGCATGGTGGATAGGTATAGAACATCCTTGGAGTACCTACAATCCACAGGCCAA
 4500 . .
 GlyValValGluAlaMetAsnHisHisLeuLysAsnGluIleSerGluThrIleValLeu
 GGAGTAGTAGAAGCAATGAATCACCATCTAAAAAACCAAATAAGTGAACAAATAGTACTA
 . .
 MetAlaIleHisCysMetAsnPheLysArgArgGlyIleGlyAspMetThrProSer
 ATGGCAATTCAATTGCATGAATTTAAAAGAAGGGGGGAATAGGGGATATGACTCCATCA
 4600 . .
 GluArgLeuIleAsnMetIleThrThrGluGlnGluIleGlnPheLeuGlnAlaLysAsn
 GAAAGATTAAATCAATATGATCACACAGAACAAAGAGATACAATTCCCTCCAAGCCAAAAT
 . .
 SerLysLeuLysAspPheArgValTyrPheArgGluGlyArgAspGlnLeuTrpLysGly
 TCAAAATAAAAGATTTCGGGTCTATTCAAGAGAAGGAGATCAGTTGTGGAAAGGA
 4700 . .
 ProGlyGluLeuLeuTrpLysGlyGluGlyAlaValLeuValLysValGlyThrAspIle
 CCTGGGAAACTACTGTGGAAAGGAGAAGGAGCAGTCCTAGTCAAGGTAGGAACAGACATA
 4800 . .
 LysIleIleProArgArgLysAlaLysIleIleArgAspTyrGlyGlyArgGlnGluMet
 MetGluGluAspLysArgTyr
 AAAATAATACCAAGAAGGAAAGCCAAGATCATCAGAGACTATGGAGGAAGACAAGAGATG
 . .
 AspSerGlySerHisLeuGluGlyAlaArgGluAspGlyGluMetAla
 IleValValProThrTrpArgValProGlyArgMetGluLysTrpHisSerLeuValLys
 GATAGTGGTCCCACCTGGAGGGTGCAGGGAGGATGGAGAAATGGCATAGCCTGTCAA
 4900 . .
 TyrLeuLysTyrLysThrLysAspLeuGluLysValCysTyrValProHisHisLysVal
 GTATCTAAAATACAAACAAAGGATCTAGAAAAGGTGTGCTATGTTCCCCACCATAAGGT
 . .
 GlyTrpAlaTrpTrpThrCysSerArgValIlePheProLeuLysGlyAsnSerHisLeu
 GGGATGGGCATGGTGGACTTGCAGCAGGGTAATATTCCCATTAAAAGGAAACAGTCATCT
 5000 . .
 GluIleGlnAlaTyrTrpAsnLeuThrProGluLysGlyTrpLeuSerSerTyrSerVal
 AGAGATAACAGGCATATTGGAACTTAAACACCAAGAAAAGGATGGCTCTCTTATTCACT
 5100 . .
 ArgIleThrTrpTyrThrGluLysPheTrpThrAspValThrProAspCysAlaAspVal
 AAGAATAACTTGGTACACAGAAAAGTTCTGGACAGATGTTACCCCAGACTGTGCAGATGT
 . .
 LeuIleHisSerThrTyrPheProCysPheThrAlaGlyGluValArgArgAlaIleArg
 CCTAATACATAGCACTTATTCCCTTGCTTACAGCAGGTGAAGTAAGAAGAGCCATCAG
 5200 . .
 GlyGluLysLeuLeuSerCysCysAsnTyrProArgAlaHisArgAlaGlnValProSer
 AGGGAAAAGTTATTGTCCTGCTGCAATTATCCCCGAGCTCATAGAGCCCAGGTACCGTC
 . .
 LeuGlnPheLeuAlaLeuValValValGlnGlnAsnAspArgProGlnArgAspSerThr
 MetThrAspProArgGluThrValPro
 ACTTCATTTCTGGCTTAGTGGTAGTGCAACAAATGACAGACCCAGAGACAGTAC
 5300 . .
 ThrArgLysGlnArgArgArgAspTyrArgArgGlyLeuArgLeuAlaLysGlnAspSer
 ProGlyAsnSerGlyGluGluThrIleGlyGluAlaPheAlaTrpLeuAsnArgThrVal
 CACCAAGGAAACAGCGCGAAGAGACTATGGAGAGGCTTCGCCCTGGCTAACAGGACAG
 5400 . .
 ArgSerHisLysGlnArgSerSerGluSerProThrProArgTyrPheProGlyVal
 GluAlaIleAsnArgGluAlaValAsnHisLeuProArgGluLeuIlePheGlnValTrp
 TAGAAGCCATAAACAGAGAAGCACTGAATCACCTACCCCCGAGAACTTATTTCAGGTGT
 (fig.1A-suite 4)

6 / 35

7/35

ThrProLeuCysValAlaMetLysCysSerSerThrGluSerSerThrGlyAsnAsnThr
AACACCTTATGTGTAGCAATGAAATGCAGCACAGAGAGCACAGGGAAACAC

ThrSerLysSerThrSerThrThrThrProThrAspGlnGluGlnGluIleSer
AACCTAAAGAGCACAAAGCACAAACCACACCCACAGACCAGGAGCAAGAGATAAG
6500

GluAspThrProCysAlaArgAlaAspAsnCysSerGlyLeuGlyGluGluThrIle
TGAGGATACTCCATGCCACCGCAGACAACACTGCTCAGGATTGGAGAGAAACGAT
6600

AsnCysGlnPheAsnMetThrGlyLeuGluArgAspLysLysGlnTyrAsnGluThr
CAATTGCCAGTTCAATATGACAGGATTAGAAAGAGATAAGAAAAACAGTATAATGAAAC

TrpTyrSerLysAspValValCysGluThrAsnAsnSerThrAsnGlnThrGlnCysTyr
ATGGTACTCAAAAGATGTGGTTGTGAGACAAATAATAGCACAAATCAGACCCAGTGTAA
6700

MetAsnHisCysAsnThrSerValIleThrGluSerCysAspLysHisTyrTrpAspAla
CATGAACCATTGCAACACATCACTCACAGAACATCATGTGACAAGCACTATTGGATGC

IleArgPheArgTyrCysAlaProProGlyTyrAlaLeuLeuArgCysAsnAspThrAsn
TATAAGGTTTAGATACTGTGCACCACCGGTTATGCCCTATTAAAGATGTAATGATAACCAA
6800

TyrSerGlyPheAlaProAsnCysSerLysValValAlaSerThrCysThrArgMetMet
TTATTCAAGGCTTGCACCCAACCTGTTCTAAAGTAGTAGCTTCTACATGCACCAAGGATGAT
6900

GluThrGlnThrSerThrTrpPheGlyPheAsnGlyThrArgAlaGluAsnArgThrTyr
GGAAACGCAAACCTCCACATGGTTGGCTTAATGGCACTAGAGCAGAGAATAGAACATA

IleTyrTrpHisGlyArgAspAsnArgThrIleIleSerLeuAsnLysTyrTyrAsnLeu
TATCTATTGGCATGGCAGAGATAATAGAACTATCATCAGCTAAACAAATATTATAATCT
7000

SerLeuHisCysLysArgProGlyAsnLysThrValLysGlnIleMetLeuMetSerGly
CAGTTGCATTGTAAGAGGCCAGGGATAAGACAGTGAAACAAATAATGCTTATGTCAGG

HisValPheHisSerHisTyrGlnProIleAsnLysArgProArgGlnAlaTrpCysTrp
ACATGTGTTCACTCCCACCTACCAGCCGATCAATAAGACCCAGACAAGCATGGTGC
7100

PheLysGlyLysTrpLysAspAlaMetGinGluValLysGluThrLeuAlaLysHisPro
GTTCAAAGGCAAATGGAAAGACGCCATGCAGGAGGTGAAGGAAACCTTGCAAAACATCC
7200

ArgTyrArgGlyThrAsnAspThrArgAsnIleSerPheAlaAlaProGlyLysGlySer
CAGGTATAGAGGAACCAATGACACAAGGAATATTAGCTTGCAGGCCAGGAAAGGCTC

AspProGluValAlaTyrMetTrpThrAsnCysArgGlyGluPheLeuTyrCysAsnMet
AGACCCAGAAGTAGCATACTGACTAACTGCAGAGGAGTTCTACTGCAACAT
7300

ThrTrpPheLeuAsnTrpIleGluAsnLysThrHisArgAsnTyrAlaProCysHisIle
GACTTGGTTCTCAATTGGATAGAGAATAAGACACACCCGCAATTATGCACCGTGCCATAT

LysGlnIleIleAsnThrTrpHisLysValGlyArgAsnValTyrLeuProProArgGlu
AAAGCAAATAATTAAACACATGGCATAAGGTAGGGAGAAATGTATATTGCTCCAGGGA
7400

GlyGluLeuSerCysAsnSerThrValThrSerIleIleAlaAsnIleAspTrpGlnAsn
AGGGGAGCTGTCCTGCAACTAACAGTAACCAGCATAATTGCTAACATTGACTGGAAAAA
7500

AsnAsnGlnThrAsnIleThrPheSerAlaGluValAlaGluLeuTyrArgLeuGluLeu
CAATAATCAGACAAACATTACCTTGTGAGAGGTGGCAGAACTATAACAGATTGGAGTT

GlyAspTyrLysLeuValGluIleThrProIleGlyPheAlaProThrLysGluLysArg
GGGAGATTATAATTGGTAGAAATAACACCAATTGGCTCGCACCTACAAAAGAAAAAAG
7600

(fig.1A-suite 6)

8/35

TyrSerSerAlaHisGlyArgHisThrArgGlyValPheValLeuGlyPheLeuGlyPhe
ATACTCCTCTGCTCACGGGAGACATAAAAGAGGTGTTCGTGCTAGGGTTCTGGGTT

LeuAlaThrAlaGlySerAlaMetGlyAlaAlaSerLeuThrValSerAlaGlnSerArg
TCTCGCAACAGCAGGTTCTGCAATGGGCAGCGTCCTGACCGTGTGGCTCAGTCCC
7700

ThrLeuLeuAlaGlyIleValGlnGlnGlnGlnLeuLeuAspValValLysArgGln
CACTTACTGGCCGGATAGTGAGCAACAGCAACAGCTGTGGACGTGGTCAAGAGACA
7800

GlnGluLeuLeuArgLeuThrValTrpGlyThrLysAsnLeuGlnAlaArgValThra
ACAAGAACTGTTGCGACTGACCGTCTGGGAACGAAAAACCTCCAGGCAAGAGTCACTGC

IleGluLysTyrLeuGlnAspGlnAlaArgLeuAsnSerTrpGlyCysAlaPheArgGln
TATAGAGAAGTACCTACAGGACCAGGCGCGCTAAATTGATGGGGATGTGCGTTAGACA
7900

ValCysHisThrThrValProTrpValAsnAspSerLeuAlaProAspTrpAspAsnMet
AGTCTGCCACACTACTGTACCATGGGTTAACATGATTCTTAGCACCTGACTGGGACAATAT

ThrTrpGlnGluTrpGluLysGlnValArgTyrLeuGluAlaAsnIleSerLysSerLeu
GACGTGGCAGGAATGCCAAAAACAAAGTCCCCTACCTGGAGGCAAATATCAGTAAAAGTT
8000

GluGlnAlaGlnIleGlnGlnGluLysAsnMetTyrGluLeuGlnLysLeuAsnSerTrp
AGAACAGGCACAAATTCAAGCAACAGAAAAATATGTATCAACTACAAAAATTAAAGCTG
8100

AspIlePheGlyAsnTrpPheAspLeuThrSerTrpValLysTyrIleGlnTyrGlyVal
GGATATTTGGCAATTGGTTGACTAACCTCCTGGGTCAAGTATATTCAATATGGAGT

LeuIleIleValAlaValIleAlaLeuArgIleValIleTyrValValGlnMetLeuSer
Val

GCTTATAATAGTAGCAGTAATAGCTTAAGAATAGTGATATATGTAGTACAAATGTTAAG
8200

AlaCysPheLeuPheProProArgLeuTyrProThrAsp
ArgLeuArgLysGlyTyrArgProValPheSerSerProProGlyTyrIleGlnGlnIle
GlyLeuGluArgAlaIleGlyLeuPheSerLeuProProProValIleSerAsnArgSer
TAGGCTTAGAAAGGGCTATAGGCCCTGTTCTCTCCCCCCCCGGTTATATCCAACAGAT

ProTyrProGlnGlyProGlyThrAlaSerGlnArgArgAsnArgArgArgArgTrpLys
HisIleHisLysAspArgGlyGlnProAlaAsnGluGluThrGluGluAspGlyGlySer
IleSerThrArgThrGlyAspSerGlnProThrLysLysGlnLysLysThrValGluAla
CCATATCCACAAGGACCGGGACAGCCACGCAAGAAACAGAAAGACGGTGGAAAG
8300

GlnArgTrpArgGlnIleLeuAlaLeuAlaAspSerIleTyrThrPheProAspProPro
AsnGlyGlyAspArgTyrTrpProTrpProIleAlaTyrIleHisPheLeuIleArgGln
ThrValGluThrAspThrGlyProGlyArg
CAACGGTGGAGACAGATACTGGCCCTGGCCGATAGCATATACATTCTGATCCGCCA
8400

AlaAspSerProLeuAspGinThrIleGlnHisLeuGlnGlyLeuThrIleGlnGluLeu
LeuIleArgLeuLeuThrArgLeuTyrSerIleCysArgAspLeuLeuSerArgSerPhe
GCTGATTGCCCTCTGACCAAGACTATACAGCATCTGCAGGGACTTACTATCCAGGAGCTT

ProAspProProThrHisLeuProGluSerGlnArgLeuAlaGluThr
LeuThrLeuGlnLeuIleTyrGlnAsnLeuArgAspTrpLeuArgLeuArgThrAlaPhe
CCTGACCCCTCCAACCTCATCTACCAAGAACTCTCAGAGACTGGCTGAGACTTAGAACAGCCTT
8500

LeuGlnTyrGlyCysGluTrpIleGlnGluAlaPheGlnAlaAlaAspAlaThrArg
MetGlyAlaSerGlySerLysLysHisSerArgProProArgGlyLeuGlnGlu
CTTGCAATATGGGTGCCAGTGGATCCAAGAACGATTCCAGGCCGCCGAGGGCTACAAG

(fig.1A-suite 7)

9/35

Glu Thr Leu Ala Gly Ala Cys Arg Gly Leu Trp Arg Val Leu Glu Arg Ile Gly Arg Gly
 Arg Leu Leu Arg Ala Arg Ala Gly Ala Cys Gly Gly Tyr Trp Asn Glu Ser Gly Gly Glu
 AGAGACTCTGGCGGGCGGCTGGAGGGTATTGGAACGAATCGGGAGGG
 8600
 Ile Leu Ala Val Pro Arg Arg Ile Arg Gln Gly Ala Glu Ile Ala Leu Leu
 Tyr Ser Arg Phe Gln Glu Gly Ser Asp Arg Glu Gln Lys Ser Pro Ser Cys Glu Gly Arg
 AATACTCGGGTTCCAAGAAGGATCAGACAGGGAGCAGAAATCGCCCTCCTGTGAGGGAC
 8700
 Gln Tyr Gln Gln Gly Asp Phe Met Asn Thr Pro Trp Lys Asp Pro Ala Ala Glu Arg Glu
 GGCAGTATCAGCAGGGAGACTTATGAATACTCCATGGAAGGACCCAGCAGCAGAAAGGG

 Lys Asn Leu Tyr Arg Gln Gln Asn Met Asp Asp Val Asp Ser Asp Asp Asp Asp Gln Val
 AGAAAAATTGTACAGGCAACAAAATATGGATGATGTAGATTGACATGATGATGACCAAG
 8800
 Arg Val Ser Val Thr Pro Lys Val Pro Leu Arg Pro Met Thr His Arg Leu Ala Ile Asp
 TAAGAGTTCTGTACACACAAAAGTACCAACTAACGACCATAGACACATAGATTGGAATAG

 Met Ser His Leu Ile Lys Thr Arg Gly Gly Leu Glu Gly Met Phe Tyr Ser Glu Arg Arg
 ATATGTCACATTTAATAAAACAAAGGGGGACTGGAAGGGATTTTACAGTGAAAGAA
 8900
 His Lys Ile Leu Asn Ile Tyr Leu Glu Lys Glu Glu Gly Ile Ile Ala Asp Trp Gln Asn
 GACATAAAATCTAAATATACTTAGAAAAGGAAGAAGGGATAATTGCAGATTGGCAGA
 9000
 Tyr Thr His Gly Pro Gly Val Arg Tyr Pro Met Phe Phe Gly Trp Leu Trp Lys Leu Val
 ACTACACTCATGGGCCAGGAGTAACATAACCAATGTTCTTGGGTGGCTATGGAAGCTAG

 Pro Val Asp Val Pro Gln Glu Gly Glu Asp Thr Glu Thr His Cys Leu Val His Pro Ala
 TACCACTAGATGTCCCACAAGAAGGGAGGACACTGAGACTCACTGCTTAGTACATCCAG
 9100
 Glu Thr Ser Lys Phe Asp Asp Pro His Gly Glu Thr Leu Val Trp Glu Phe Asp Pro Leu
 CACAAACAAAGCAAGTTTGATGACCCGCATGGGAGACACTAGTCAGGGAGTTGATCCCT

 Leu Ala Tyr Ser Tyr Glu Ala Phe Ile Arg Tyr Pro Glu Glu Phe Gly His Lys Ser Gly
 TGCTGGCTTATAGTTACGAGGCTTTATTGGTACCCAGAGGAATTGGGCACAAGTCAG
 9200
 Leu Pro Glu Glu Trp Lys Ala Arg Leu Lys Ala Arg Gly Ile Pro Phe Ser
 GCCTGCCAGAGGAAGAGTGGAAAGGGAGACTGAAAGCAACAGGAATACCATTAGTTAAA
 9300
 GACAGGAACAGCTATACTTGGTCAGGGCAGGAAGTAACAAACAGAAACAGCTGAGACTGC

 AGGGACTTCCAGAAGGGCTGTAACCAAGGGAGGGACATGGGAGGAGCTGGTGGGAAC
 9400
 GCCCTCATATTCTCTGTATAAAATACCCGCTAGCTTGCATTGTACTTCGGTCGCTCTGC

 GGAGAGGCAGATTGAGGCCCTGGGAGGTTCTCCAGCAGTAGCAGGTAGAGCCTGG
 9500
 GTGTTCCCTGCTAGACTCTCACCAAGCACTTGGCCGGTGTGGCAGACGGCCCCACGCTT
 9600
 GCTTGCTTAAAAACCTCCTTAATAAGCTGCCAGTTAGAAGCA

(fig.1A-suite 8)

10/35

FIG 1B

AGTCGGCTCTGGGAGAGGCTGGCAGATTGAGCCCTGGAGGTCTCTCCAGCACTAGCAG
 GTAGAGCCTGGGTGTTCCCTGCTAGACTCTCACCCACACTGGCCGGTGTGGGAGACT
 GGCTCCACGCTTGCTTAAAGACCTCTCAATAAAAGCTGCCATTTAGAAGTAAGCTA
 100
 GTGTGTGTTCCCATCTCTCTAGTCGGCGCTGGTCAACTCGGTACTCGGTAAATAAAAG
 200
 ACCCTGGCTGTTAGGACCCCTGGCTGTTAGGACCCCTTGCTTGGAAACCGAAGCA
 300
 GGAAAATCCCTAGCAGATTGGCGCCCGAACAGGGACTGAAGGAGACTGAGAGACTCCTG
 AGTACGGCTGAGTGAAGGCACTAAGGGCGCAGGAACCAACCACGACGGAGTGCTCCTAG
 400
 AAAGGCGCGGGTGGTACAGACGGCGTGAGGAGCGGGAGAGAAGAGGCTCCTGGTTG
 CAGGTAACTGCAACACAAAAAGGAATAGCTGTCTTATCCAGGAAGGGATAATAAGAT
 500
 GAGDMETGLYALARGASN SERVALLEUSERGLYLYSALASPGLULEUGLU
 AGAGTGGGAGATGGCGCAGAAACTCCGTCTTCAGGGAAAGAACAGATGAATTAGA
 600
 LYSILEARGLEUARGPROGLYGLYLYSLSLYSTYRMETLEULYSHISVALVALTRPALA
 AAAAATTAGACTACGACCCGGCGAAAGAAAAGTACATGTTGAAGCATGTAGTAGTATGGC
 ALAASNGLULEUASPARGPHEGLYLEUALAGLUSERLEULEUGLUASNLYSGLUGLYCYS
 AGCAATGAATTAGATAGATTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATG
 700
 GLNLYSILEUSERVALLEUALAPROLEUVALPROTHRLYSERGLUASNLUYSSER
 TCAAAAAATACTTCCGTCTAGCTCATTAGTCCAACAGGCTCAGAAAATTAAAAAG
 LEUTYRASNTHRVALCYSVALILETRPCYSILEHISALAGLUGLULYSVALYSHISTR
 CCTTTATAACTGTCTGCGTCATCTGGTCATTCAACAGGAGAAAGTGAACACACAC
 800
 GLUGLUALALYSGLNILEVALGLNARHGHSLEUVALMETGLUTHRLYTHRALAGLUTHR
 TGAGGAAGCAAACAGATAGTCAGAGACACCTAGTGTGAAACAGGAACAGCAGAAAC
 900
 METPROLYSTHRSERARGPROTHRALAPROPHESERGLYARGGLYGLYASNTYRPROVAL
 TATGCCAAAAACAAGTAGACCAACAGCACCATTAGCGGCAGAGGAGGAAATTACCCAGT
 GLNGLNILEGLYGLYASNTYRTHRHSLEUPROLEUSERPROARGTHRLEUASNALATRP
 ACAACAAATAGGGTAACTATACCCACCTACCATTAAGCCCCGAGAACATTAAATGCC
 1000
 VALLYSLUEILEGLUGLULYSLYSPHEGLYALAGLUVALVALSERGLYPHEGLNLALEU
 GGTAAAATTAAATAGGGAGAAATTGGAGCAGAAGTAGTGTCAAGGATTTCAGGCACT
 SERGLUGLYCYSLEUPROTYRASPILAEASNGLNMETLEUASNCYSVALGLYASPHISGLN
 GTCAGAAGGCTGCCCTCCATGACATTAATCAGATGTTAAATTGTGGAGACCATCA
 1100
 ALAALAMETGLNILEILEARGASPILIEASNGLUGLUALALAASPTRPASPLEUGLN
 AGCGGCTATGCAGATCATCAGAGATATTATAATGAGGAGGCTCCAGATTGGACTTGCA
 1200
 HISPROGLNGLNALAPROGLNGLNLYGLNLEUARGGLUPROSERGLYSERASPILEALA
 GCACCCACAACAGCTCCACACAAGGACAGCTTAGGGAGCCGTCAAGGATCAGATATTGC
 GLYTHRTHRHSERTHRVALGLUGLUGLNILEGLNTRPMETTYRARGGLNGLNASNPOLIE
 AGGAACAACAGTACAGTAGAAGAACAAATCCAGTGGATGTACAGACAAACAGAACCCAT
 1300

FIG. 1B-

11/35

PROVALGLYASNILETYRARGARGTRPILEGLNLEUGLYLEUGLNLSCYSVALARGMET
 ACCAGTAGGCAACATTACAGGAGATGGATCAAATGGGTGCAAAATGTGTAGAAT
 TYRASNPROTHASNILEEUASPVALLYSGLNGLYPROLYSGLUPROPHEGLNSERTYR
 GTATAACCCAACAAACATTCTAGATGAAAACAAGGCCAAAAGGCCATTTCAGAGCTA
 1400
 VALASPARGPHETYRLYSSERLEUARGALAGLUGLNTHRASPPOALAVALLYSASNTRP
 TGTAGACAGGTTCTACAAAAGTTAACAGCAGAACACAGATCCAGCAGTAAAGAATTG
 1500
 METTHRGLNTHRLEULEUILEGGLNASNALAASNPROASPCYSLSLEUVALLEULYSGLY
 GATGACTCAAACACTGCTGATTCAAATGCTAACCCAGATTGCAAGCTAGTGCCTGAAGGG
 LEUGLYTHRASNPROTHRLEUGLUGLUMETLEUTHRALACYSGLNGLYVALGLYGLYPRO
 GCTGGGTACGAATCCCACCCCTAGAAGAAATGCTGACGGCCTGCAAGGAGTAGGGGGCC
 1600
 GLYGLNLYSALAARGLEUMETALAGLUALALEULYSGLUAIALEUALAPROALAPROILE
 AGGACAGAAGGCTAGATTAATGGCAGAACAGGCCCTGAAAGAGGCCCTCGCACCCAGGCCAAT
 * * * * * POLVALLEUGLULEUTRP
 PROPHEALAAALAGLNLNLYSGLYPROARGLYSPOITLEYSCYSTRPASNCYSGLY
 CCCTTTGCAGGCCAACAGAACAGGCCAAAGAAAGCCATTAAAGTGTGAAATTGTGG
 1700
 GLUGLYARGTHRLEUCYSLYSALAMETGLN SERPROLYSLYSTHRLYMETLEUGLUMET
 LYSGLUGLYHISSERALAARGGLNCYSARGALAPROARGARGGLNLGYCYSTRPLYSCYS
 GAAGGAAGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAACAGACAGOGATGCTGAAATG
 1800
 TRPLYSASNGLYPROCYSTYRGLYGLNMETPROLYSLNTHRGLYGLYPHEPHEARGPRO
 GLYLYSHETASPHISVALMETALALYS CYS PROASNARGGLN ALA GLY PHELEUGLYLEU
 TGGAAAAATGGACCATGTTATGGCAAATGCCAAACAGACAGGCCCTTGTAGGCT
 TRPPROLEUGLYLYSGLUALAPROGLNPHEPROHISGLYSER SERALASERGLYALAASP
 GLYPROTRPGLYLYSLSYSPROARGASNPHEPROMETALAGLNV AL HIS GLNGLYLEUHR
 TGGCCCTTGGGAAAGAACAGGCCCTGAAATTTCCCATGGCTCAAGTCATCAGGGCTGAC
 1900
 ALAASNCYS SERPRODARGARGTHR SERCYSGLYSER ALA LYSGLUEU HIS SALA LEUGLY
 PROTHR ALAPROPROGLUGLUPROAL AVALASPLEULEULYSASNTYRMETHISLEUGLY
 GCCAACTGCTCCCCAGAACAGACAGCTGTGGATCTGCTAAAGAAACTACATGCACTTGG
 GLNALAALAGLUARGLYSGLNARGGLUALALEUGLNLGLYGLYASPARGLY PHEALALA
 LYSGLNLNARGGLUSERARGGLYLYSPROTYRLYSGLUVALTHRGLUASPLEULEU HIS
 CAAGCAGCAGAGAACAGAGGGAAAGCAGAGGCCCTACAAGGAGGTGACAGAGGATTGCTGCA
 2000
 PROGLNPHE SERLEUTRPARGARGPROVALYALTHR ALAHISILEGLUGLYGLNPROVAL
 LEUASNSERLEUPHEGLYGLYASPGLN
 CCTCAATTCTCTCTTGGAGGAGACAGTAGTCACTGCTCATATTGAAGGACAGCCTGTA
 2100
 GLUVALLEULEUASPTHRGLYALAAASPASP SERILEVALTHR GLYILEGLULEUGLYPRO
 GAAGTATTATTAGATAACAGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCA
 HISTYRTHRPROLYSILEYALGLYGLYILEGLYGLYPHEILEASNTHR LYSGLUTYRLYS
 CATTATAACCCAAAAATAGTAGGAGGAATAGGAGGTTTATTAAACTAAAGAACATACAAA
 2200
 ASNVALGLUILEGLUVALLEUGLYLYSARGILELYSGLYTHRILEMETTHR GLYASPTHR
 AATGTAGAAATAGAAGTTTACAGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCA
 PROILEASNILEPHEGLYARGASNL EULEUTHR ALALEUGLYMETSERLEU ASNL EUPRO
 CCGATTAACATTGGTAGAAATTACTAACAGCTCTGGGATGTCTCAAATCTTCCC
 2300
 ILEALALYSVALCLUPROV ALLYS SERPROLEULYS PROGLYLYSASPGLYPROLYS EU
 ATAGCTAAGGTAGGCCCTGAAAGTCGCCCTAAAGCCAGGAAGGATGACAGGGACACC
 2400
 LYSGLNTRPPROLEUSERLYSGLULYSILEVALA EALEUARGGLUITLECYSGLULYSMET
 AACAGTGCCATTATCAAAGAAAAGATAGTTGCTTAAGAGAAATCTGTGAAAAGATG

(Fig. 1B-suite 1)

12/35

GLULYSASPLGLNLEUGLUGLUALAPROPROTHRASNPROTYRASNTHRPROTHRPH
 GAAAAAGATGGTCAGTGGAGGAAGCTCCCCCACCAATCCATAACACCCCCACATT
 2500
 ALAILELYSLYSASPLYSASNLYSTRPARGMETLEUILEASPPHEARGLULEUASN
 GCTATAAACAAAAGGATAAAAACAAATGGAGAATGCTGATAGATTAGGGAACTAAAT
 ARGVALTHRGLNASPPHETHRGLUVALGLNLEUCLYILEPROHISPROALAGLYLEUALA
 AGGGTCACTCAAGACTTACGGAAAGTCCAATTAGGAATACCACACCCCTGCAGGACTACCA
 2600
 LYSARGLYSARGILETHRYALLEUASPILEGLYASPALATYRPHESERILEPROLEUASP
 AAAAGGAAAAGGATTACAGTACTGGATATAGGTGACGCATATTCTCTACCTCTAGAT
 2700
 GLUGLUPHEARGGLNTYRTHRALAPHETHRLEUPROSERVALASNASHNALAGLUPROGLY
 GAAGAATTAGGCAGTACACTGCCTTACTTTACCATCAGTAAATAATGCAGAGCCAGGA
 LYSARGTYRILETYRLYSALLEUPROGLNGLYTRPLYSGLYSERPROALAILEPHEGLN
 AAACGATACATTATAAGGTTCTGCTCAGGGATCGAAGGGTCACCAGCCATCTTCAA
 2800
 TYRTHRMETARGHISVALLEUGLUPROPHEARGLYSALAASNPROASPVALTHRLEUVAL
 TACACTATGAGACATGTGCTAGAACCTTCAGGAAGGAAATCCAGATGTGACCTTAGTC
 GLNTYRMETASPASPILELEUILEALASERASPARGLTHRASPLEUGLUHISASPARGLY
 CAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACAGGGTA
 2900
 VALLEUGLNLEULYSGLULEULEUASNTERILEGLYPHESERSERPROGLUGLULYSPE
 GTTTACAGTAAAAGAACCTTTAAATAGCATAGGGTTTCATCCCCAGAAGAGAAAATTC
 3000
 GLNLYSASPPROPROPHEGLNTRPMETGLYTYRGLULEUTRPPROTHRPLYSTRPLYSLEU
 CAAAAAGATCCCCATTCAATGGATGGGTACGAATTGTGGCCGACAAATGGAAGTTG
 GLNLYSILEGLULEUPROGLNARGGLUTHRTRPVALASNASPILEGNLYSLEUVAL
 CAAAGATAGAGTTGCCACAAGAGAGACCTGGACAGTGAAATGATATAACAGAAGTTAGTA
 3100
 GLYVALLEUASNTRPALAALAGLNILETYRPROGLYILELYSTHRLYSHISLEUCYSARG
 GGAGTATTAATTGGGCAGCTCAAATTATCCAGGTATAAAACCAAAACATCTCTGTAGG
 LEUILEARGLYLYSMETTHRLEUTHRGLUGLUVAGLNTRPTHRGLUMETALAGLUALA
 TTAATTAGAGGAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAGAAGCA
 3200
 GLUTYRGLUGLUASNLYSILEILELEUSERGLNGLUGLUGLYCYSTRTYRGLNGLU
 GAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTTACCAAGAA
 3300
 SERLYSPROLEUGLUALATHRVALILELYSSERGLNAPASGLNTRPSERTYRLYSILE
 AGCAAGCCATTAGAACGCCACGGTGATAAAAGAGTCAGGACAATCAGTGGCTTATAAAATT
 HISGLNGLUASPLYSILEULYSVALGLYLYSPHEALALYSILELYSASNTHRHI
 CACCAAGAACAAAATCTGAAAGTAGGAAATTGCAAAGATAAAAGAACATACACATACC
 3400
 ASNGLYVALARGLEUULEUALAHISVALILEGLNLYSILEGLYLYSGLUALAILEVALILE
 AATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGAACATAGTGTAC
 TRPGLYGLNVALPROLYSPHEHISLEUPROVALGLULYSASPYVALTRPGLUGLNRPTP
 TGGGACAGGTCACCAACTTACAGTTGAGAAGGATGATGGAACAGTGGTGG
 3500
 THRASPTYRTRPGLNVALTHRTRPILEPROGLUTRPASPPHEILESERTHRPROPROLEU
 ACAGACTATTGGCAGGTAACCTGGATACCGGAATGGATTTCATCTAACACACCATTAA
 3600
 VALARGLEUVALPHEASNLEUVALLYSASPPROFILEGLUGLYGLUGLUTHRTYRTYRVAL
 GTAAGATTAGTCTCAATCTAGTGAAAGGACCTATAGAGGGAGAACCTATTATGTA
 ASPGLYSERCYSSERLYSGLNTERLYSGLUGLYLYSALAGLYTYRILETHRASPARGGLY
 GATGGATCATGTAGTAAACAGTCAAAAGAACAGGAAAGCAGGATATCACAGACAGGGC

(fig.1B-suite 2)

13/35

3700
 LYSASPLYSVALLYSVALLEUCLUGLNTHRTHRASNGLNGLNALAGLULEUGLUALAPHE
 AAAGACAAGGTAAAAGTGTAGAACAGACTACTAATCAACAAGCAGAAATTGGAAGCATT
 LEUMETALALEUTHRASP SERGLYPROLYSALAASNILEILEVALASP SERCLNTYRAL
 CTCATGGCATTGACAGACTCAGGCCAAAGGCAAATTATAGTAGACTCACAATATGTT
 3800
 METGLYILEILETHRGLYCYSPROTHRGLUSERGLUSERARGLEUVALASNGLNILEILE
 ATCGGAATAAACAGGATGCCCTACAGAACAGAGCAGGCTAGTTAACCAAAATAATA
 3900
 GLUGLUMETILELYSLYSTHRGLUILEYTYRALALATRPVALPROALAHISLYSGLYILE
 GAAGAAATGATCAAAAGACAGAAATTATGTGGCATGGTACCCAGCACACAAAGGTATA
 GLYGLYASNGLNGLUILEASPHISLEUVALSERGLNGLYILEARGGLNYVALLEUPHELEU
 GGAGGAACCAAGAAATAGACCACCTAGTTAGTCAGGGATTAGACAAGTTCTCTTCTG
 4000
 GLULYSILEGLUPROALAGLNGLUHISERLYSTYRHISERASNILELYSGLULEU
 GAAAGATAGGCCAGCACAGAACATAGTAACATACCATAGTAACATAAAAGAATTG
 VALPHELYSPHEGLYLEUPROARGLEUVALALALYSGLNILEVALASP THRCYSAPLVS
 GTATTCAAATTGGATTACCCAGACTAGTGCCAAACAGATAGTAGACACATGTGATAAA
 4100
 CYSHISGLNLYSGLYGLUALAILEHISGLYGLNVALASNERSPLEUGLYTHRTRPGLN
 TGTATCAAAAGGAGAAGCTACATGGCAGGTAATTCAAGACCTAGGGACTTGC
 4200
 METASPCYSTHRHISLEUGLUGLYLYSILEVALILEVALALALVALHISVALALASERGLY
 ATGGATTGTACCCATCTAGAGGGAAAAATAGTCATAGTTGCAGTACATGTAGCTAGTGA
 PHEILEGLUALAGLUVALILEPROGLNGLUTHRGLYARGGLNTHRALALEUPHELEU
 TTCATAGAACAGAAGTAATTCCACAAAGAACAGGAAGACAGACAGCACTATTCTGTTA
 4300
 LYSLEUALASERARGTRP PROILETHR HISLEUHISTHRASPASNGLYALAASNPHEA
 AAATTGGCAAGCAGATGCCATTACACATCTGCACACAGATAATGGTCTAACTTGCT
 SERGLNGLUVALLYSMETVÁLALATRPTPRPALAGLYILEGLUHISTRPHEGLYVALPRO
 TCGCAAGAACTAAAGATGGTGCATGGTGGCAGGGATAGGGCACACCTTGGGTACCA
 4400
 TYRASNPROCLNSERGLNGLYVALVALGLUALAME TASNHISHISLEULYSASNGLNILE
 TACAATCCACAGACTCAGGGAGTAGTGGAAAGCAATGAATCACCACCTGAAAAATCAAATA
 4500
 ASPARGILEARGGLUGLNALAASN SERVALGLUTHRILEVAL EUMETALAYALHIS CYS
 GATAGAACATCAGGAACAAGCAAATTCAAGTACATGAAACCATAGTATTATGGCAGTTCATG
 METASNPHELYSARGARGGLYGLYILEGLYASPMETTHR PROALAGLUARGLEU
 ATGAATTAAAGAAGGGAGGAATAGGGATATGACTCCAGCAGAAAGATTAATTAAAC
 4600
 METILETHRTHRGLUGLNGLUILEGGLNPHEGLNLSERLYSASNSERLYSPHELYSASH
 ATGACTACAGAACAGAAATACAATTCAACAATCAAAACTCAAAATTAAAAAT
 PHEARGVALTYRTYRARGGLUGLYARGASPGLNLEUTRPLYSGLYPROGLYGLULEU
 TTTCGGGTCTATTACAGAGAACGGCAGAGATCAGCTGTGGAAAGGGACCCGGTAGCTATTG
 4700
 TRPLYSGLYGLUGLYALAVALILEEULYSVALGLYTHRASPILELYSVALVALPROARG
 TGAAAGGGGAAGGAGGAGTCATCTAAAGGTAGAACAGACATTAAGGTAGTACCCAGG
 4800
 ARGLYSALALYSILEILELYSASPTYRGLYGLYLYSGLUMETASPSERSERSEHIS
 QHETGLUGLUGLULYSARGTRP ILEVALVALPROTHR
 AGAAAGGCTAAATTATCAAAGATTATGGAGGAGAACAGATGGATAGTAGTTCCCAC
 METGLUASPTHRGLYGLUALAARGGLUVALALA
 TRPARGILEPROGLUARGLEUGLUARGTRPHIS SERLEUYLELYSTYRLEULYSTYRLYS
 ATGGAGGATACGGAGAGGCTAGAGAGGTGGCATAGCCTCATAAAATATTGAAATATAA
 4900
 (fig. 1B - suite 3)

THRLYSASPLEUGLNLYSALACYSTYRAL PROHISIISLYSVAL GLYTRPALATRPTRP
 AACTAAAGATCTACAAAAGGCTTGCTATGTGCCCATCATAAGCTGGATGGGCATGGT
 THRCYSSERARGVALILEPHEPROLEUGLNLGLUGLYSERHISLEUGLUVALGLNGLTYR
 GACCTGCAGCAGAGTAATCTCCCACACAGGAAGGAAGCCATTAGAAGTACAAGGTA
 5000
 TRPASNLEUTHRPROGLUARGGLYTRPLEUSERTHRTYRALAVALARGILETHRPTYR
 TTGGAATTTGACACCAGAAAGACGGCTGCTCAGTACTTATGCACTGAGGATAACCTGGT
 5100
 SERLYSASPPHETRPTHRASPVALTHRPROGLUTYRALAASPILELEULEUHISSETRH
 CTCAAAGGACTTTGGACAGATGTAACACCAGAAATATGCAGATACTTACTGCATAGCAC
 TYRPHEPROCYSPHETHRALAGLYGLUVALARGARGALAILEARGGLYGLUARGLEU
 TTATTTCCCTTGCTTACAGCGGGAGAAGTGAGAAGGGCCATCAGGGAGAACGACTGCT
 5200
 SERCYSCYSARGPHEPRDARGALAHISLYSHISGLNLVALPROSERLEUGLNTYRLEUALA
 GTCTTGCTGCAGGTCCCAGAGCTCATAGCACCAGGTACCAAGTCTACAGTACTTAGC
 LEUARGVALVALSERHISVALARGSERGLNGLYGLUASNPROTHRTRPLSGLNTRPARG
 X METSERASPPROARGGLUARGILEPROPROGLYASNSERGLYGLU
 ACTGAGAGTAGTCACTGTCAGATCCCAGGGAGAATCCACCTGAAACAGTGGAG
 5300
 ARGASPASNARGSERLEUARGVALALALYSGLNASNSERARGGLYASPLYSGLNARG
 GLUTHRILEGLYGLUVALAPHEGLUTRPLEUASNARGTHRVALGLUGLUILEASNARGGLU
 AAGAGACAATAGGAGAACCGCTTCAGTGGCTAACACAGAAGTAGAGGAGATAAACAGAG
 5400
 GLYGLYLYS PROPROTHRGLUGLYALAASNPHPROGLYLEUALALYSVALLEUGLYILE
 ALAVALASNHSILEUPROARGGLULEUILEPHEGLNVALTRPGLNARGSERTRPGLUTYR
 AGGCCTGAAACCACTACCGAGGGAGCTAATTTCAGGTTGGCAAGGCTTGGAAAT
 LEUALA
 TRPHISASPGUGLNLGLYMETSERGLNSERTYRTHRLYSTYRARGTYRLEUCYSLEUILE
 ACTGGCATGATGAACAAGGGATGTCACAAAGCTATACAAAATACAGATACTTGTGTTAA
 5500
 GLNLYSALALEUPHEMETHISCSLYSLYSGLYCYSARGCYSLYGLUGLYHISGLY
 TACAAAAGGCTTATTATGCATTGCAAGAAAGGCTGTAGATGTCAGGGAGAACAG
 ALAGLYGLYTRPARGPROGLYPROPROPROPROPROPROGLYEUALA R METGLU
 GGGCAGGGGGATGGAGACCAGGACCTCCTCTCCCCCTCAGGACTAGCATAATGG
 5600
 GLUARGPROPROGLUASNLUGLYPROGLNARGGLUPROTTPASPGULTRPYALVALGLU
 AAGAAAGCCTCCAGAAAATGAAGGCCACAAAGGGAACCATGGATGAGTGGTAGTG
 5700
 VALLEULYSGLULEULYSGLUGLUALALEULYSHISPHEASPPROARGLEULEUTHRALA
 AAGTTCTGAAAGAACTGAAAGAAGCTTAAACCATTTGATCTCGGCTCTAACCG
 TATI METGLUTHRPROLEUARGGLUGLNLGLUASN
 LEUGLYASHNHSILETYRASNARGHISGLYASPTHRLEUGLUGLYALAGLYGLULEUILE
 CACTTGTAATCATATCTATAATAGACATGGAGACACCCTGAGGAGCAGGAGAACTCA
 5800
 LEUGLUSERSERASNGLUARGSERERTYRILESERGLUALAALALAALAILEPROGLU
 ARGILELEUGLNLARGALALEUPHEILEHISPEARGSERGLYCYSSERHISSETRARGILE
 TTGAAATCCTCAACGAGCGCTTCTACATTCAGAACGGCTGCAGCCATTCCAGAA
 SERALAASNLEUGLYGLUGLULILELEUSERGLNLLEUTYRARGPROLEUGLUALACYSTYR
 GLYGLNPROGLYGLYGLYASNPROLEUSERTHRILEPROPROSERARGSERMETLEU
 TCGGCCAACCTGGGGAGGAATCCTCTCTCAACTATACCGCCCTCTAGAACCATGCTAT
 5900
 ASNTHRCYSTYRCYSLYSLYSCYSTYRHISCSYSLNPHECYSPELEULYSLYSGLY
 AACACATGCTATTGCAAAAGTGTGCTACCATTGCCAGTTTTCTAAAAAGGGC
 6000
 LEUGLYILESERTYRGLULYSSERHISARGARGARGTHRPROLYSLYSALALYSALA
 ARTIMETARGSERHISTHRGLYGLUGLUGLULEUARGARGCARGLEU

(fig.1B-suite 4)

15/35

TTGGGGATAACTTATCAGAACTCACACAGGAGAAGAAGAACTCGAAGAAGCTAACCC
 ASNTHRSERSERALASERASHGLU
 ILEHISLEULEUHISGLNTHRSERLYSTYRGLYLEUSERTRPLYSSERALALAATYRARG
 ENV MET GLYCYSLEUGLYASNLNEULEUILEALA
 AATACATCTTCTGCATCAAACCGAGTAAGTATCGGGTCTGGAAATCACCTGGCTTATCG
 6100
 HISLEULEU
 ILECYSSERLYSCYSLEUTRPILEILECYSILECLNTYRVALTHRVALPHETYRGLYVAL
 CCATCTGCTCAAGTGTCTATGGATTATTTGATTCAATATGTACAGTCAGTCTTTATGGTG
 PRODALATRPARGASHNALATHRILEPROLEUPHECYSLAATHRLYSASNARGASPTHRTRP
 TACCAAGCTTGGAGGAATGCCACAATTCCCCCTCTTCTGTCACCAACAAAGATAGGGATACTT
 6200
 GLYTHRTHRGLNCYSLEUPRGASPASNAPSPTYRSEERGLULEUALALEUASNVALTHR
 GCGAACAACTCAGTGCCTACAGATAATGATGATTATTCAGAAITGGCCCTTAATGTTA
 6300
 GLUSERPHEASPALATRPGLUASNTHRVALTHRGLUGLNALALEGLUASPVATRPGLN
 CAGAAAGCTTGATGCTTGGAGAAATACAGTCACAGAACAGGAATAGAGGACGTATGGC
 LEUPHEGLUTHRSERILELYSPROCYSVALLYSLEUSERPROLEUCYSILETHRMETARG
 AACTCTTGAGACCTCAATAAAGCCTTGTAATTCACCAATTGCAATTACTATGA
 6400
 CYSASNLYSSERGLUTHRASPLYSTRPGLYLEUTHRLYSERERTHRTHRTHRALASER
 GATGCAATAAAAGTGAAGACAGATAATGGGATTGACAAAATCATCA&CAACACAGCAT
 6500
 THRTHRTHRTHRTHRALALYSSERVALGLUTHRARGASPILEVALASNGLUTHRSER
 CAACAACAACAACACAGCAAATCAGTAGAGACAAGAGACATAGTCATGAGACTA
 PROCYSVALVALHISASPNSCYSTHRLYLEUGLUGLNGLUPROMETILESERCYSLYS
 GTCCCTTGCTAGTTGATGATAATTGCACAGGCTTGGAAACAAGGCCAATGATAAGCTGTA
 6600
 PHEASNMETTHRGLYLEULYSARGASPLYSLYSGLUTYRASNLGLUTHRTRPTYRSER
 AATTCAACATGACAGGGTAAAAAGAGACAAGAAAAGAGGACTAACATGAAACTTGGTACT
 ALAASPLEUVALCYSGLUGLNGLYASNNSERTHRLYASNLUSERARGCYSTYRMETASN
 CTGCAGACTGGTTGTGAACAAGGAATAGCACTGGTAATGAAAGTAGATGTTACATGA
 6700
 HIS CYS ASNTHR SERVAL ILE GLN GLUCYSCYSASPVYRTRPASPALEI ARG
 ATCACTGTAATACTCTGTTATCCAAGAGTGTGTGACAAAGATTATGGGATGCTTATA
 CYSARGTYRCYSLAPROPROGLYTYRALALEUEUARGCYSASNPSPTHRA SNTYRSER
 GATGTAGATATTGTGCACCTCCAGGTTATGCTTGTAGTGAATGACACAAATTATT
 6800
 GLYPHENETPROASN CYSSERLYSVALVALVALSERSERCYSTHRARGNETGLUTHR
 CAGGCTTATGCCTAACTGTTCAAGGTAGTGGTCCTCATGCACAGGATGATGGAGA
 6900
 GLNTHRSERTHRTRPHEARGPHEASNLGYTHRARGCALAGLUASNARGTHR TYRILETYR
 CACAGACTTCACTGGTTCCGTTAATGCAACTAGAGCAGAAAATAGAACCTATATT
 TRPHISGLYARGASPASNARGTHRILEILESERLEUASNLYSHISTYRASNLLEUTHRNET
 ACTGGCATGGTAGAGATAATAGGACTATAATTAGCTAAATAAGCATTATAATCTAACAA
 7000
 LYS CYSARGARGPROGLYASHNLYSTHRYALLEUROVALTHRILEMETSERALALEUVAL
 TGAAATGTAAGACCAGGAATAAGACAGTTTACCACTGTCACCAATTATGTCGCTTGG
 PHEHISERGLNPROVALASNGLUARGPROLYSGLNHALATRPCYSARGPHEGLYGLYASN
 TTTCCACTCACAAACCAAGTCATGAGAGGCCAAAGCAGGCATGGTAGGTTGGAGAA
 7100
 TRPLYSGLUALAILELYSGLUVALLYSGLNTHRILELEVALLYSHISPROARGTYRTHRLY
 ATTGGAAAGGAGGCAATAAAAGAGGTGAAGCAGACCAATTGTCAAACATCCCAGGTACTG
 7200
 THRASNTHRASPLYSILEASNLLEUTHRALAPROARGGLYLYASPPROGLUVALTHR
 GAACAAACAACTGATAAAATCAATTGACGGCTCCTAGAGGAGGAGATCCGGAAAGTTA

(fig.1B-suite 5)

PHE MET YRP THR AS NC Y S ARG GLY GLU PHE LEU TY R C Y S L Y S M E T A S N T R P P H E L E U A S N
 CCT T C A T G T G G A C A A A T T G C A G A G G A G T T T C T C T A C T G T A A A T G A A T T G G T T C T A A
 7300
 TRP V A L G L U A S P A R G S E R L E U T H R T H R C L N L Y S P R O L Y S G L U A R G H I S L Y S A R G A S N T Y R
 A T T C C G T A G A A C A T A G G A G T C T A A C T A C C C A G A A G G C A A A G G C A T A A A G G A A T T
 7400
 V A L P R O C Y S H I S I L E A R G G L N I L E I L E A S N T H R T R P H I S L Y S V A L G L Y L Y S A S N V A L T Y R
 A C G T A C C A T G T C A T A T T A G A C A A A T A C A A C A C T T G G C A T A A A G T A G G C A A A A T G T T T
 7500
 L E U P R O P R O A R G G L U G L Y A S P L E U T H R C Y S A S N S E R T H R V A L T H R S E R L E U I L E A L A A S N
 A T T G C C T C C A A G A G A G G G G A G A C C T C A C G T G T A A C T C C A C A G T G A C C A G T C T C A T A G C A A C
 7600
 I L E A S N T R P T H R A S P G L Y A S N G L N T H R S E R I L E T H R M E T S E R A L A G L U V A L A L A G L U L E U
 A C A T A A A T T G G A C T G A T G G A A C C A A A C T A G T A T C A C C A T G A G T G C A G A G G T G G C A G A A C
 T Y R A R G L E U G L U L E U G L Y A S P T Y R L Y S L E U V A L G L U I L E T H R P R O I L E G L Y L E U A L A P R O
 T G T A T C G A T T G G A A T T G G G A G A T T A A A T T A G T A G A A T C A C T C C A A T T G G C T T G G C C
 7700
 T H R A S N V A L L Y S A R G T Y R T H R T H R G L Y G L Y T H R S E R A G S N L Y S A R G G L Y V A L P H E V A L
 C C A C A A A T G T C A A G A G G T A C A T C A T G G T G G C A C C T C A A G A A A T A A A G A G G G G T C T T G
 7800
 L E U G L Y P H E L E U G L Y P H E L E U A L A T H R A L A G L Y S E R A L A M E T G L Y A L A A L A S E R L E U T H R
 T C C T A C G G T T C T T G G G T T T C T C G C A A C G G C A G G T T C T G C A A T G G C G C G G C G T C G T T G A
 7900
 V A L T H R A L A G L N S E R A R G T H R L E U L E U A L A G L Y I L E V A L G L N G L N G L N G L N L E U L E U
 C C G T G A C C G C T C A T C C C G G A C T T T A T T G G C T G G G A T A G T G C A G C A A C A G C A A C A G C T G T
 8000
 A S P V A L V A L L Y S A R G G L N G L N G L U L E U L E U A R G L E U T H R V A L T R P G L Y T H R L Y S A S N L E U
 T C C A G C T G G T C A A G A G A C A A C A A G A A T T G T T G C G A C T G A C C G T C T G G G G A A C A A A G A A C C
 8100
 G L N T H R A R G V A L S E R A L A I L E G L U L Y S T Y R L E U L Y S A S P G L N A L A G L N L E U A S N A L A T R P
 T C C A G A C T A G G G T C T C T G C C A T C G A G A A G T A C T T A A G G A C C A G G G C C A G C T A A A T G C T T
 8200
 A S N I L E T H R A L A L E U G L U G L U A L A G L N I L E G L N G L N G L U L Y S A S N M E T T Y R G L U L E U
 C A A A T A A C G G C C T C C T A G A A G A G G C A C A A A T T C A C A A G A G A A C A T G T A G A A T
 8300
 G L N L Y S L E U A S N S E R T R P A S P V A L P H E G L Y A S N T R P P H E A S P L E U T H R S E R T R P I L E L Y S
 T A C A A A A G T T G A A T A G C T G G G A T G T G T T G G C A A T T G T T G A C C T T A C T T C T G G A T A A
 T Y R I L E G L N T Y R G L Y I L E T Y R I L E I L E V A L G L Y V A L I L E L E U A R G I L E V A L I L E T Y R
 A G T A T A C A A T A T G G A A T T T A T A T A A T T G T A G G A G T A A T A C T G T T A A G A A T A G T G A T C T
 8400
 I L E V A L G L N M E T T L E U A L A A R G L E U A R G G L N G L Y T Y R A R G P R O V A L P H E S E R S E R P R O P R O
 A T A T A G T A C A A A T G C T A G C T A G G T T A A G A C A G G G G T A T A G G C C A G T G T T C T C T C C C C A C
 T A T 2 A R C P R O I L E P R O A S N A R G I L E A R G L E U C Y S C L N P R O L Y S L Y S A L A
 A R T 2 V A L A S P P R O T Y R P R O T H R G L Y S E R G L Y S E R A L A S N G L N A R G A R G G L N
 S E R T Y R P H E G L N * * * T H R H I S T H R G L N G L N A S P P R O A L A E U P R O T H R L Y S G L U G L Y L Y S
 C C T C T T A T T C C A G T A G A C C C A T A C C C A A C A G G A T C C G G C T G G C C A A C C A A A G A A G G C A
 L Y S L Y S G L U T H R V A L G L U A L A A L A V A L A L A T H R A L A P R O G L Y L E U G L Y A R G * T A T (f in)
 L Y S A R G A R G A R G T R P A R G G L N A R G T R P G L N G L N L E U L E U A L A E U A L A A S P A R G I L E T Y R
 L Y S G L Y A S P G L Y G L Y G L Y S E R G L Y G L Y A S N S E R S E R T R P P R O T R P G L N I L E G L U T Y R I L E
 A A A A A G G G A C A C G G T G G A G G C A G C G G T G G C A A C A G G T C C T G G C C T T G G C A G A T A G A A T A
 (fig. 1B-suite 6)

17/35

SERPHEPROASPPROPROTHRASPTHRPROLEUASPLEUALATLEGLNGLNCEUGLNASH
 HISPHELEUILEARGGLNLEUILEARGLEULEUTHRTRPLEUPHESERASNCYSARGTHR
 TTCATTTCCTGATCCCCAACTGTACCCCTCTTGACTTGGCTATTCAAGCAACTGCAGAA
 LEUALAILEGLUSERILEPROASPPROPROTHRASNILEPROGLUALAEUCYSALEU
 LEULEUSERARGALATYRGLNILELEUGLNPROILEPHECLNARGLEUSERALATHRTYR
 CCTTGCTATCGAGAGCATACCAGATCCTCAACCAATTCCAGAGGCTCTGCGACCT
 8500 F MET GLY GLY ALA
 ARG ARG ILE ARG ARG SER PRO GLN ALA • ART2 (fin)
 GLY GLU PHE GLY GLU VAL LEU ARG LEU GLU LUE UTH RTY RLEU GLN TYR GLY TRP SERT YR
 ACG GAG AATT CCG GAG AAG T C CT CAG GCT TGA ACT GAC C TAC C TAC A AT AT GGG T GG AG CT
 ILE SER LY SLYS ARG SER LY S PRO PRO GLU ILE CYS AS P ARG ASP SER CYS GLY ARG VAL
 PHE CLN GLU ALA VAL GLN ALA ALA A RG A S PLE U AR GL N AR GL E U LE U AR GL A AR G GL Y
 ATT T C CA AGA AAC C CG GT C CA AG C CG CC AG A G AT CT G CG A Ç AG A G ACT C TT G C G G G C G T G
 8600
 GLY ARG AS NT Y R GL Y ARG LEU PHE LY S GL Y V AL GLU A SP GL Y S ER SER GL N S ER LE U GL Y
 GL U LY S LE U TR P GL U AL A LE U GL N AR G GL Y GL Y AR G TR P I LE L E U AL A I LE P RO A RG A RG
 GGG AGA AA ATT AT GGG AGG CT T C AA AGG CT GGA AG AT GG AT C T C T C G C A AT C C T C A G G A
 8700
 GL Y LEU A S PLY S GL Y LEU S ER SER LEU S ER C Y S GL U GL Y GL N LY STY RA S N GL N GL Y GL U
 ILE ARG GL Y GL Y LEU GL U LE U TH RLE U LEU •
 GG AT TAG AC A AGG C TT GAG C TACT C TT GAG G CC AAA A AT ACA AT CAGG AG A A
 TYR MET AS N H R P O T R P A R G A S N P D A L A G L U G L U A R G L Y S L Y S L E U P R O T Y R A R G L Y S
 TAC AT GA A TA C T C C A T G G A G A A C C C A G C T G A A G A G A G G A A A A A T T A C C A T A C A G A A A A
 8800
 GL N A S N I L E A S P A S P I L E A S P G L U G L U A S P A S P A S P L E U V A L G L Y I L E P R O V A L G L U A
 C A A A A T A G A T G A T A T A G A T G A G G A A G A T G A T G A C T T G C T A G G G A T A C C A G T T G A G G C C
 AR G Y A L P R O L E U A R G T H R M E T S E R T Y R L Y S L E U A L A I L E A S P M E T S E R H I S P H E I L E L Y S
 A C A G T T C C C C T A A G A A C A A T G A G T T A C A A A T T G G C A A T A G A T A T G T C T C A T T T T A T A A A A
 8900
 GL U LY S GL Y GL Y LEU GL U GL Y I L E I L E P R O A S P T R P G L N I L E H I S S E R G L Y P R O G L Y I L E
 G A A A A G G G G G A C T G G A A G G C A T T A T T A C A G T G C A A G A A G C A T A G A A T C T T A G A C A T A
 9000
 T Y R L E U G L U L Y S G L U G L U G L Y I L E I L E P R O A S P T R P G L N I L E H I S S E R G L Y P R O G L Y I L E
 T A C T T A G A A A G G A A G G C A T C A T C C A G A T T G G C A G A T A C A C T C C G G A C C A G G A A T T
 A R G T Y R L E U L Y S M E T P H E G L Y T R P L E U T R P L Y S L E U I L E P R O V A L A S N V A L S E R A S P G L U
 A G A T A C C T A A A G A T G T T G G C T G G C T A T G G A A A T T A T C C C T G T A A A T G T A T C A G A T G A G
 9100
 A L A G L N G L U A S P G L U C L U H I S T Y R L E U V A L H I S P R O A L G L N T H R S E R G L N T R P A S P A S P
 G C A C A G G A G G A T G A G G A C T T A T T A G T G C A C C C A G C T C A A A C T C C C A G T G G G A T G A C
 P R O T R P G L Y G L U V A L L E U A L A T R P L Y S P H E A S P P R O T H R L E U A L A T Y R T H R T Y R G L U A L A
 C C T T G G G G A G A G G T T C T A G C A T G G A A G T T T G A T C C A A C T C T A G G C T A C A C T T A T G A G G C A
 9200
 T Y R I L E A R G T Y R P R O G L U L P H E G L Y S E R L Y S S E R G L Y L E U S E R G L U L Y S G L U V A L L Y S
 T A T A T T A G A T A C C C A G A A G A G T T T G G A A G C A A G T C A G G C T G T C A G A G A A A G G A G G T T A A A
 9300
 A R G A R G L E U A L A A R G G L Y L E U L E U G L U M E T A L A A S P A R G L Y S G L U T H R S E R
 A G A A G G C T A G C C C A A G G A G G C T T C T G G A A A T G G C T G A C C G A A G G A A A C T A G C T G A G A C
 A G C A G G G A C T T C C A C A A G G G A T G T C A T G G G A G G T A C T G G G A G G A G C C G G T T G G G A A
 9400
 C A C C C A C T T T C T G A T G T A T A A A T A T C A T G C A T T C G C T C T G T A T T C A G T C G C T C T G C G
 G A G A G G C T G G C A G A T T G A G C C C T G G G A G G T T C T C C A G C A C T A G C A G G T A G A G C C T G G G
 9500
 T G T T C C C T G C T A G A C T C T C A C C A G C A C T T G G C C G G T G C T G G G C A G A G T G G C T C C A C G C T T
 9600

(fig.1B-suite 7)

18/35

FIG. 1C

séquence LTR
 CIVET
 versus
 HIV-2 ROD

| | | | | | | |
|---|------|------|------|------|------|------|
| X | 8960 | 8970 | 8980 | 8990 | 9000 | 9010 |
| TGGAAGGGATTATTACAGTGCAGAACATAGAAATCTTAGACATATACTTAGAAAAGG | | | | | | |
| ::: ::::: :: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: | | | | | | |
| TGGAAGGGATGTTTACAGTGAAGAACATAAAATCTTAAATATATACTTAGAAAAGG | | | | | | |
| X | 8950 | 8960 | 8970 | 8980 | 8990 | |
| | | | | | | |
| | 9020 | 9030 | 9040 | 9050 | 9060 | |
| AAGAAGGCATCATACCAGATTGGCAGATACTCCGGA---CCAGGAATTAGATACCTAA | | | | | | |
| ::: ::::: :: :: ::::: ::::: :: : ::::: ::::: ::::: ::::: ::::: | | | | | | |
| AAGAAGGGATAATTGCAGATTGGCAGAACTACACTCATGGGCCAGGAGTAAGATAACCAA | | | | | | |
| 9010 | 9020 | 9030 | 9040 | 9050 | | |
| | | | | | | |
| | 9080 | 9090 | 9100 | 9110 | 9120 | |
| AGATGTTGGCTGGCTATGAAATTAAATCCCTGTAAATGTATCAGATGAGGCACAGGAGG | | | | | | |
| ::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: | | | | | | |
| TGTTCTTGGCTGGCTATGAAAGCTAGTACCACTAGATGTCCCACAAGAAGGGGAGGACA | | | | | | |
| 9070 | 9080 | 9090 | 9100 | 9110 | | |
| | | | | | | |
| | 9140 | 9150 | 9160 | 9170 | 9180 | |
| ATGAGGAGCATTATTAGTCACCCAGCTAAACTTCCCAGTGGATGACCCCTGGGAG | | | | | | |
| ::: ::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: | | | | | | |
| CTGAGACTCACTGCTTAGTACATCCAGCACAAACAAGCAAGTTGATGACCCGATGGG | | | | | | |
| 9130 | 9140 | 9150 | 9160 | 9170 | | |
| | | | | | | |
| | 9200 | 9210 | 9220 | 9230 | 9240 | |
| AGGTTCTAGCATGGAAGTTGATCCAACCTAGCCTACACTTATGAGGCATATATTAGAT | | | | | | |
| ::: ::::: :: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: | | | | | | |
| AGACACTAGTCTGGAGTTGATCCCTTGCTGGTTATAGTTACGAGGCTTTATCGGT | | | | | | |
| 9190 | 9200 | 9210 | 9220 | 9230 | | |
| | | | | | | |
| | 9260 | 9270 | 9280 | 9290 | 9300 | |
| ACCCAGAAAGAGTTGGAAGCAAGTCAGGCCTGTCAGAGAAAGAGGTTAAAGAAGGCTAG | | | | | | |
| ::: ::::: :: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: | | | | | | |
| ACCCAGAGGAATTGGGCACAAGTCAGGCCTGCCAGAGGAAGAGTGGAAAGGCAGACTGA | | | | | | |
| 9250 | 9260 | 9270 | 9280 | 9290 | | |
| | | | | | | |
| | 9320 | 9330 | 9340 | 9350 | | |
| CCGCAAGAGGCCCTTGTAAATGGCT-GACAGGAAGGAAACT----- | | | | | | |
| ::: ::::: :: : :: ::::: ::::: :: | | | | | | |
| AAGCAAGAGGAATACCATTAGTAAAGACAGGAACAGCTATACTTGGTCAGGGCAGGAA | | | | | | |
| 9310 | 9320 | 9330 | 9340 | 9350 | | |

FIG. 1C

19 / 35

9360 9370 9380 9390
 ----- AGCTGAGACAGCAGGGACTTCCACAAGGGATGTCATG--GGGA
 ::::::::::::::::::::: ::::::: :::: : :::
 GTAACATAACAGAAACAGCTGAGACTGCAGGGACTTCCAGAAGGGCTGTAACCAAGGCA
 9370 9380 9390 9400 9410
 9400 9410 9420 9430 9440 9450
 GGTACTGGGGAGGAGCCGGTGGAACACCCACTTCTTGATGTATAATATCACTGGAT
 :: :: ::::::: :::: ::::::: :::: : : ::::::: :::: : :::
 GGGACATGGGAGGAGCTGGTGGGAACGCCCTCATATTCTGTGATGATAAAATACCCGCTA
 9430 9440 9450 9460 9470
 9460 XX 10 20 30 40
 TTTCGCTCTGTA--TTCTGGAAGGGATTTATTACAGTGCAAGAAGACATAGAATCTTAGAC
 : ::::::::::::: : ::::::: ::::::: ::::::: ::::::: :::
 GCTTGCATTGTACTTCTGGAAGGGATGTTTACAGTGAAAGAAGACATAAAATCTTAAAT
 9490 XX 10 20 30 40
 50 60 70 80 90
 ATATACTTAGAAAAGGAAGAAGGCATCATACCAGATTGGCAGATAACTCCGGA---CCA
 ::::::::::::::::::::: ::::: :: ::::::: ::::::: :::: : :::
 ATATACTTAGAAAAGGAAGAAGGATAATTGCAGATTGGCAGAACTACACTCATGGGCCA
 50 60 70 80 90 100
 110 120 130 140 150
 GGAATTAGATACTAAAGATGTTGGCTGGCTATGGAAATTAAATCCCTGTAATGTATCA
 :: :: ::::::: :: : : :: ::::::: ::::::: :: : :: :: :::
 GGAGTAAGATAACCAATGTTCTTGGGTGGCTATGGAAAGCTAGTACCAAGTAGATGTCCCCA
 110 120 130 140 150 160
 170 180 190 200 210
 GATGAGGCACAGGGAGGATGAGGAGCATTATTTAGTGCACCCAGCTCAAACCTCCAGTGG
 : :: : ::::: ::::: :: : ::::::: :: ::::::: :::: : :::
 CAAGAAGGGGAGGACACTGAGACTCACTGCTTAGTACATCCAGCACAAACAAGCAAGTT
 170 180 190 200 210 220
 230 240 250 260 270
 GATGACCCCTGGGGAGAGGTTCTAGCATGGAAGTTGATCCAACCTCTAGCCTACACTTAT
 ::::::: :: ::: ::::: ::: ::::::: :: :: :: :: :::
 GATGACCCCGCATGGGGAGACACTAGTCTGGAGTTGATCCCTGCTGGCTTATAGTTAC
 230 240 250 260 270 280
 290 300 310
 GAGGCATATATTAGATAACCCAGAAGAGTTGGAAGCA
 ::::::: :: :: :::
 GAGGCTTTATTCCGG
 290

(fig.1C-suite 1)

20/35

FIG. 2
 (HIV-2.P
 versus
 (HIV-1.P

| | | | | | | | | | | | |
|--------------|--|-----|-----|-----|----|--|--|----------|------|------|--|
| KIV2----- | | | | | | | | | | env4 | |
| | 10 | 20 | 30 | 40 | 50 | | | | | | |
| | KIQQLLIA ILLA-SACLV YC-CYVTVFY GVPYKHNATI | | | | | | | | | | |
| HIV1----- | HRVKEKYQHL WRWCWKWGTN LLGILMICSA TEKLWVTVYY GVPVWKEATT | | | | | | | | | | |
| 5 | | | | | | | | | | | |
| HIV2----- | | | | | | | | | env5 | 100 | |
| | 60 | 70 | 80 | 90 | | | | | | | |
| | PLFCATRNR -DT-- WG T1QCLPDND YOEITL-NVT EAFAWNNTV | | | | | | | | | | |
| HIV1----- | TLFCASDAKA YDTEVENVWA THACVPTDPN PQEVVLVHVT ENFNMWKNDM | | | | | | | | | | |
| 10 | | | | | | | | | | | |
| KIV2----- | | | | | | | | | env6 | 150 | |
| | 110 | 120 | 130 | 140 | | | | | | | |
| | TEQAIEDVWH LFETSIKPCV KLTPLCVAMKICSSSTESTGN NTTSKSTSTT | | | | | | | | | | |
| HIV1----- | VEQMHHEDIIS LWDQSLKPCV KLTPLCVSLK CTDL---GN ATNTHSSNTN | | | | | | | | | | |
| 15 | | | | | | | | | | | |
| HIV2----- | | | | | | | | | 160 | 200 | |
| | 170 | 180 | 190 | 200 | | | | | | | |
| | -TTTPTDQE QEISEDTPCA RADNCGLGE SETINCQFNH TGLERDKKKQ | | | | | | | | | | |
| HIV1----- | SSSGEMHHHEK GEIK --- HCSFNIS TSIRGKVQKE YAFFYKLDII | | | | | | | | | | |
| 20 | | | | | | | | | | | |
| HIV2----- | | | | | | | | 210 | 250 | | |
| | 220 | 230 | 240 | 250 | | | | | | | |
| | Y--NET-WYS KVYCETNNST NQTQCYMNHC HTSVITESCD KHYWDALRER | | | | | | | | | | |
| HIV1----- | PIDNDTTSYT --- TSC HTSVITQACP KVSFEPIPIH | | | | | | | | | | |
| 25 | | | | | | | | | | | |
| HIV2----- | | | | | | | | 260 env8 | 300 | | |
| | 270 | 280 | 290 | 300 | | | | | | | |
| | YCAPPGYALL RC-NDT-WYS GFAPNC SKVV ASTCTR NMET QTSTWF-GFN | | | | | | | | | | |
| HIV1----- | YCAPAGFAIL KCNNKTFNGT GP---CTNVS TVQCTHGIRP VVSTQLLL-N | | | | | | | | | | |
| 30 | | | | | | | | | | | |
| KIV2----- | | | | | | | | 310 | 350 | | |
| | 320 | 330 | 340 | 350 | | | | | | | |
| | GTRAE---H RTIYWHGRD H-RTII-SLN KYYNLSLECK RPGNKTVKQI | | | | | | | | | | |
| HIV1----- | GSLAEEEVVI RSANFT---D NAKTII VQLN QSVE--INCT RPNNNTRKSI | | | | | | | | | | |
| 35 | | | | | | | | | | | |
| HIV2----- | | | | | | | | 360 | 400 | | |
| | 370 | 380 | 390 | 400 | | | | | | | |
| | MLMS--GHVF HS HYQPINKR PROAWCFKG -RUKDANQEV KETLAKHPRY | | | | | | | | | | |
| HIV1----- | RIQRGPGEAF VTIGKIGN-- MRQAHCHISR AKWJAT---L KQIASKLREQ | | | | | | | | | | |

FIG. 2

21 / 35

↓

| | | | | | | |
|-------|-----------|--|---------------|--------------------|---------------|-----|
| 5 | HIV2----- | 410 RGTNDTRNIS FAAPGKGSDP EVAYMWTNCR GEFLYCNMTW FLH--WI | 420 * * * * * | env10 430 ** * * * | 440 * * * * | 450 |
| | HIV1----- | FGHHKT-II FKQSS-GGDP EIVTHSFNCG GEFFYCNSTQ LFNSTWFNST | | | | |
| | | | | | | |
| 10 | HIV2----- | 460 EN KTHRNAYAPCH IKOIINTWHK VGRNVYLPPR EGELSCNSTV | 470 * * * * * | env11 480 ** * * * | 490 * * * * | 500 |
| | HIV1----- | WSTEGLNNTE GSDTITLPCR IKQFINNWQE VGKAMYAPPI SGQIRCSSNI | | | | |
| | | | | | | |
| 15 | HIV2----- | 510 TSIIANIDWQ NNNQTNITFS AEVAELYRL- ELGDYKLV EITPIGFAPT | 520 * * * * * | 530 ** * * * | 540 * * * * * | 550 |
| | HIV1----- | TGLLLTRDGQ NNNNGSEIFR PGCGDHRDNW RSELYKYKVV KIEPLGVAPT | | | | |
| | | | | | | |
| 20 | HIV2----- | env3 560 KEKRYSSAHG RHTRGVFVLG FLGFLATA GSAMGAAS LTVSAQSRTL | 570 * * * * * | 580 ***** | 590 * * * * * | 600 |
| | HIV1----- | KAKRR--VVQ. REKRAVGI-G ALFLGFLGAA GSTNGARSMT LTVQA--RQL | | | | |
| | | | | | | |
| 25 | HIV2----- | 610 LAGIVQQQQQ LLDVVVKRQQE LLRLTVWGTK NLQARVTAIE KYLQDQARLN | 620 * * * * * | 630 ** * * * | 640 env1 650 | |
| | HIV1----- | LSGIVQQQNN LLRAIEAQHQ LLQLTVWGIK QLQARILAVE RYLKQDQQLLC | | | | |
| | | | | | | |
| 30 | HIV2----- | 660 SWGCCAFRQVC HTTVPW VNDSLAPDWD NMWTWQEWEKQ VRYLEANISK | 670 * * * * * | 680 * * * * * | 690 * * * * * | 700 |
| | HIV1----- | IWGCSGKLIC TTAVPKHASL SNKSLEQIWH XHTWMEDRE INHYTSLIHS | | | | |
| | | | | | | |
| 35 | HIV2----- | 710 env2. 720 SLEQAOQIQQE KNMYELOKLN SDIFGNWFD LTSWVKYIQY GVLIIVAVIA | 730 * * * * * | 740 * * * * * | 750 * * * * * | |
| | HIV1----- | LIEESQNQQE KNEQELLELD KWASLWNWFH ITNWLWYIKI FIMIVGGLYVG | | | | |

(fig.2 - suite 1)

22 / 35

| | | | | | | |
|-------|-----------|------------|------------|-------------|------------|--------------------|
| | | 760 | 770 | 780 | 790 | 800 |
| | HIV2----- | LRIVIYVVQM | LSRLRKGYRP | V-FSSPPGYI | QQIHIHKDRG | QPANEETEED |
| | | *** * * | *** * | * | ** | * ** |
| | HIV1----- | LRIVFAVLSI | VNRVRQGYSP | LSFQT | — | HLPTPRG PDRPEGIEEE |
| | | | | | | |
| 5 | | 810 | 820 | 830 | 840 | 850 |
| | HIV2----- | GGSNGGDRYW | PWPIAYIRFL | IRQLIRLLT | — | LYSIC RDLLSRFLT |
| | | ** ** | ** | ** | — | **** |
| | HIV1----- | GGERDRDRSI | RLVNGSLA-L | IWDDLRLSLCL | FSYHRL | — RDLLLIVTRI |
| | | | | | | |
| 10 | | 860 | 870 | 880 | 890 | 900 |
| | HIV2----- | LQLIYQNLRD | WLRLRTA-F | LQYGCEWIQE | AFQ | AAA RATRETL |
| | | * | * | *** | — | * |
| | HIV1----- | VELLG--RRG | WEALKYWWNL | LQYWSQELKN | SAVSLLNATA | IAVAEGTDRV |
| | | | | | | |
| 15 | | 910 | 920 | 930 | 938 | |
| | HIV2----- | —AGACRG | LWRVLERIGR | GILAVPRRIK | QGAEIALL | |
| | | ** | — | ***** | ** * ** | |
| | HIV1----- | IEVVQGACRA | — | -IRHIPRRIK | QGLERILL | |
| | | | | | | |

(fig. 2 - suite 2)

23/35

(ENV-mac
FIG. 3 (versus
 (ENV-ROD

| | | | | | |
|--|-------|---------------|---------------|---------------|-------|
| 10 | 20 | 30 | 40 | 50 | |
| MGCLGNOLLIAIC--SKCLWIICIQQVTVFYGVPAWRNATIPLFCATKNRDTWGTTOCL | | | | | |
| : | ::::: | ::: | ::::::::::::: | ::::::::::::: | ::::: |
| MM----NOLLIAILLASACLVY-CTQVVTVFYGYPTWKNATIPLFCATRNRTWGTIQLC | | | | | |
| 10 | 20 | 30 | 40 | 50 | |
| 60 | 70 | 80 | 90 | 100 | 110 |
| PDNODDYSELALNVTESFOAVENTVTEQAIEDVWQLFETSIKPCVKLSPLCITMRCNKSET | | | | | |
| ::::: | ::::: | ::::: | ::::::::::::: | ::::::::::::: | ::: |
| PDNODDYQEITLNVTEAFDAWNNTVTEQAIEDVWHLFETSIKPCVKLTPLCVAMKCSSTES | | | | | |
| 60 | 70 | 80 | 90 | 100 | 110 |
| 120 | 130 | 140 | 150 | 160 | 170 |
| DKHGLTKSSTTTASTTTTAKSVETRDIVNETS--PCVVHDNCTGLEQEPMISCKFNM | | | | | |
| ::: | ::::: | ::: | ::: | ::: | ::: |
| STGNNTTSKST--STTTTTP----T-DQEQEISEDTPCARADNCSGLGEEETINCQFNM | | | | | |
| 120 | 130 | 140 | 150 | 160 | |
| 180 | 190 | 200 | 210 | 220 | 230 |
| TGLKRDKKKEYNETWYSADLVCEQGNSTGNESRCYMHNCNTSVIQECCDKDYNDAIRCRY | | | | | |
| ::: | ::::: | ::: | ::: | ::::::::::::: | ::::: |
| TGLERDKKKQYNETWYSKDVVCETNNST-NATQCYMHNCNTSVITESOKHYWDAIRFRY | | | | | |
| 170 | 180 | 190 | 200 | 210 | 220 |
| 240 | 250 | 260 | 270 | 280 | 290 |
| CAPPGYALLRCNDTNYSGFMPNC SKYVVSSCTRMMETQTSTHWFRFNGTRAENRTYIYWHG | | | | | |
| ::::::::::::: | ::::: | ::::::::::::: | ::::::::::::: | ::::::::::::: | ::: |
| CAPPGYALLRCNDTNYSGFAPNC SKVAVASTCTRMMETQTSTWFGFNGTRAENRTYIYWHG | | | | | |
| 230 | 240 | 250 | 260 | 270 | 280 |
| 300 | 310 | 320 | 330 | 340 | 350 |
| RDNRTIISLNKHYNLTMKCRRPGNKTLPVTIMSLVFHS--QPVNERPQAWCRFGGNW | | | | | |
| ::::::::::::: | ::: | ::::::::::::: | ::: | ::: | ::: |
| RDNRTIISLNKYNNLSLHCKRPGNKTVKQIMLMMSGHVHFHYQPINKRPROAWCHFKGKW | | | | | |
| 290 | 300 | 310 | 320 | 330 | 340 |
| 360 | 370 | 380 | 390 | 400 | |
| KEAIKEVKQTIVKHPRYTGTNNNTDKINLTAPRGG-DPEVTFMWTNCRGEFLYCKMNWFLN | | | | | |
| :: | ::::: | ::::: | ::: | ::::: | ::::: |
| KDAMQEVKETLAKHPRYRGTDTRNISFAAPGKGSDPEVAYMWTNCRGEFLYCNHTWFLN | | | | | |
| 350 | 360 | 370 | 380 | 390 | 400 |

FIG. 3

24/35

420 430 440 450 460
 HVEDRSLLTOKPKERHKRNYVPCHIROIINTWHKVGKNVYLPPREGDLTCNSTVTSLIAN
 :
 WIEN-----KT-H-RNYAPCHIKOINTWHKVGCRNVYLPPREGELSCNSTVTSLIAN
 410 420 430 440 450

 480 490 500 510 520
 INWTDGNQTSITHSAEVAELYRLELGOKLVEITPIGLAPTNVKRYTTG-GTSRNKRGVF
 :
 IDWONNNNOTNITFSAEVAELYRLELGOKLVEITPIGFAPTKEKRYSSAHG--RHTRGVF
 460 470 480 490 500 510

 540 550 560 570 580
 VLCFLGFLATAGSAMGAASLTVAOSRTLLAGIVQQQQQLLOVVKRQELLRLTVHGTKN
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::: :::::
 VLCFLGFLATAGSAMGAASLTVAOSRTLLAGIVQQQQQLLOVVKRQELLRLTVHGTKN
 520 530 540 550 560 570

 600 610 620 630 640
 LQTRYSAIEKYLKDOAQLNAWGCARQYCHTTYPWPNASLTPDWNNNETHQEWERKVDFLE
 :
 LQARYTAIEKYLDQDQARLNWSGCAFQYCHTTYPWYNDLAPDWONMTHQEWEKOVRYLE
 580 590 600 610 620 630

 660 670 680 690 700
 ANITALLEEAOIQQEKNMYELOKLNSWDYFGNWFDLTSWIKYI0YGIYIIIVGVILLRIVI
 :
 ANISKSLAQAOIQQEKNMYELOKLNSWDIFCNWFDLTSWIKYI0YGVLIIVAVIALRIVI
 640 650 660 670 680 690

 720 730 740 750 760
 YIVQMLARLRQYRPVFSSPPSYFQ*THTQQDPALPTKEGKKGDGGSGGNSSWPHQIEY
 :
 YYYQMLSLRKGYRPVFSSPPGYIQQIHDKDRGQPANEETEEDGGSNGCDRYWPWPIAY
 700 710 720 730 740 750

 780 790 800 810 820
 IHFLIROLIRLLTLFSNCRTLLSRAYQILQPIFORLSATYGEFGEVRLRELTYLOYGWS
 :
 IHFLIROLIRLLTRLYSICRDLLSRSLTLOLIYONLRDW-----LRLRTAFLQYGCE
 760 770 780 790 800

 840 850 860 870 880
 YFOEAQVAA-RDLRORLLRA-RGEKLWEALQRGGRWILAIPRRIROGLELTL
 :
 WIQEAFQAAAARATRETLAGACRG--LWRVLERIGRGILAVPERRIROGAEIALL
 810 820 830 840 850

(fig. 3-suite 1)

25/35

FIG. 4 (GAG-mac
 versus
 (GAG-ROD

| | | | | | |
|---|-----|-----|-----|-----|-----|
| 10 | 20 | 30 | 40 | 50 | |
| YOHKKEIAVFYPGRDNKIEHEMGARNSVLSGKKADELEKIRLRPGGKKYMLKHVVWAAN | | | | | |
| | | | | | |
| MGARNSVLRGKKADELERIRLRPGGKKYRLKHIVWAAN | | | | | |
| 10 | 20 | 30 | | | |
| | | | | | |
| 70 | 80 | 90 | 100 | 110 | |
| ELDRFGLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIAEAEKVKHTEE | | | | | |
| | | | | | |
| KLDLRFGLAESLLESKEGCQKILTVLDPMVPTGSENLKSLFNTVCVIWCIAEAEKVKDTEG | | | | | |
| 40 | 50 | 60 | 70 | 80 | 90 |
| | | | | | |
| 130 | 140 | 150 | 160 | 170 | |
| AKQIVORHLVMETGTAETMPKTSRPTAPFSGRGGNYPVQQIGGNYTHLPLSPRTLNAWVK | | | | | |
| | | | | | |
| AKQIVRRHLVAETGTAEKMPSTSRTAPSSEKGGNYPVQHYGGNYTHIPLSPRTLNAWVK | | | | | |
| 100 | 110 | 120 | 130 | 140 | 150 |
| | | | | | |
| 190 | 200 | 210 | 220 | 230 | |
| LIEEKKGAEVYSGFQALSEGCLPYDINQMLNCVGDHQAMOIRDIINEEAADHDLQHP | | | | | |
| | | | | | |
| LYVEEKKGAEVYVPGFQALSEGCTPYDINOMLNCVGDHQAMOIREIINEAAEWDVQHP | | | | | |
| 160 | 170 | 180 | 190 | 200 | 210 |
| | | | | | |
| 250 | 260 | 270 | 280 | 290 | |
| QQAPQQ-GQLREPSGSDIAGTTSTYEEQIQWMYRQQNPPIPVGNIYRRWIQLGLQKCVRMY | | | | | |
| | | | | | |
| IPGPLPAGQLREPRGSDIAGTTSTYEEQIQWMFRPQNPVPGNIYRRWIQICLQKCVRMY | | | | | |
| 220 | 230 | 240 | 250 | 260 | 270 |
| | | | | | |
| 300 | 310 | 320 | 330 | 340 | 350 |
| NPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTOPAVKNWMTQTLIQQNANPDCKLVLKGL | | | | | |
| | | | | | |
| NPTNILDIKQGPKEPFQSYVDRFYKSLRAEQTOPAVKNWMTQLLVQNANPDCKLVLKGL | | | | | |
| 280 | 290 | 300 | 310 | 320 | 330 |
| | | | | | |
| 360 | 370 | 380 | 390 | 400 | 410 |
| GTNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPAPIFAAAQQKGPRKPIKCHNCGK | | | | | |
| | | | | | |
| GMNPTLEEMLTACQGVGGPGQKARLMAEALKEVIGPAPIFAAAQQ---RKAFCWNCGK | | | | | |
| 340 | 350 | 360 | 370 | 380 | 390 |

FIG. 4

26 / 35

(fig.4 - suite 1)

27 / 35

FIG. 5 (POL-mac
 (versus
 (POL-ROD

FIG. 5

28/35

410 420 430 440 450 460
 DRYVLQLKELLNSIGFSSPEEKFQKOPPFQWMGYELWPTKWKLOKIELPORETWTYNDIQ
 :::::::::::::: :::: : ::::::: ::::::::::::::: :::: : :::::::
 DRYVLQLKELLNGLGFPDEKFQKOPPYHWMGYELWPTKWKLOK[QLPOKEINTVNDIQ
 390 400 410 420 430 440

 470 480 490 500 510 520
 KLYGVLNWAACIYPGIKTKHLTRLRKGKMTLTEEVWHTEMAEAEYEENKIILSQEDEGGY
 ::::::::::: ::::::::::::::::::::: ::::::: :::: : :::::::
 KLYGVLNWAACOLYPGIKTKHLTRLRKGKMTLTEEVWHTELAEALEENRILSQEDEGGY
 450 460 470 480 490 500

 530 540 550 560 570 580
 YOESKPLEATVIKSODNWSYKIHQEDKILKYGKFAKIKNTHTNGVRLLAHVIOKICKEA
 :::: : ::::: : : : ::::: ::::::: ::::: ::::::: ::::: :::::::
 YQEKELEATVQKDQENQWHTYKIHQEEKILKVGKYAKVKNTHTNGIRLLAQVYQKIGKEA
 510 520 530 540 550 560

 590 600 610 620 630 640
 IVIHWGQVPKFHLPVKEQWHEQWKTDYQVTWIPFISTPPLYRFLVQDPIEGEET
 :::: : ::::::: ::::::: ::::::: ::::: ::::::: ::::: :::::
 LVIHWGRPKFHLPVREIEWQWHQDNYWQVTWIPDWFVSTPPLYRFLAFNLVQDPIGAET
 570 580 590 600 610 620

 650 660 670 680 690 700
 YYVDGSCSKOSKEKGAGYITDRGKDKVVKLEQTTNQQAELEAFMALTDSPKANIIVDS
 :::: : ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: :::::::
 FYTDGSCNRQSKEKGAGYVTDRGKDKVKKLEQTTNQQAELEAFANALTDSPKVNIIIVDS
 630 640 650 660 670 680

 710 720 730 740 750 760
 QYVMGIITGCPTESESRLVNOIEEMIKKTEIYVAVPAHKGIGGNQEIIDLHVSOGIQV
 ::::: : ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: :::::::
 QYVMGISAQTESESKIYVNOIEEMIKKEAIYVAVPAHKGIGGNQEVIDLHVSOGIQV
 690 700 710 720 730 740

 770 780 790 800 810 820
 LFLEKIEPAQEEHSKYHSNIKELVFKGLPRLVAKQIYDTCOKCHQKGEAIHGQVNSOLG
 ::::::::::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::::::::::
 LFLEKIEPAQEEHEKYHSNVKELSHKFGIPNLVARQIYNSCAOQQKGEAIHGQVNAELG
 750 760 770 780 790 800

 830 840 850 860 870 880
 TWQMDCTHLEGKIVIVAVHYASGFIAEVIPQESTGROTALFLKLASRWPTJHLHTDNGA
 ::::::::::::: ::::::::::::: ::::::::::::: ::::::::::::: :::::::::::::
 TWQMDCTHLEGKIIIVAVHYASGFIAEVIPQESTGROTALFLKLASRWPTJHLHTDNGA
 810 820 830 840 850 860

 890 900 910 920 930 940
 NFASQEVKMVAHHWAGIEHTFGVPPNQPSQGVVEAMNHHLKNCIDRIREQANSVETIVLMA
 :: ::::::: ::::: ::::::::::::: ::::::::::::: ::::::: :::::::
 NFTSQEVKMVAHHWIGEQSFGVPPNQPSQGVVEAMNHHLKNCISRIREQANTIEIVLMA
 870 880 890 900 910 920

 950 960 970 980 990
 VHCMNFKRRGGIGDMTPAERLINMITTEOEIQFOQSKNSKFKNFRVYYREGRDOLWKPG
 ::::::::::::: ::::::::::::: ::::::: ::::::: :::::::::::::
 IHCMNFKRRGGIGDMTPSERLINMITTEOEIQFOLOAKNSKLKDFRVYFREGRDOLWKPG
 930 940 950 960 970 980

 1010 1020 1030 1040 1050
 ELLWKCEGAVILKVCTDIKVYPRRKAKIIKOYGGKEMDSSSHMETGEAREYA

(fig.5-suite 1)

29/35

::::::: ::::::: ::::::: ::::: :: : : :
VELLKGEGLVKGTDIKIIPRRKAKIIRDYGGRQEMDSGSHLEGAREDGEHA
990 1000 1010 1020 1030

(fig. 5-suite 2)

30/35

FIG. 6
 (Q.mac
 (versus
 (Q.ROD

| | | | | |
|--|-----|-----|-----|-----|
| 10 | 20 | 30 | 40 | 50 |
| MEEEKRHIVVPTWRIPERLERWHSLIKYLKYTKDLQXACYYPHHKVGHAWHTCSRVIFF | | | | |
| ::: :::::::::: : : : :::::: :::::::::::: : ::::::::::::::::::::: | | | | |
| MEEDKRHIVVPTWRVPGRMKWHSLVKYLKYTKDLEKVCYVPHHKVGHAWHTCSRVIFF | | | | |
| 10 | 20 | 30 | 40 | 50 |
| | | | | |
| 70 | 80 | 90 | 100 | 110 |
| LOEGSHLEYQGYHNLTPERGWLSTYAVRITWYSKDFWTDTPEYADILLHSTYFPCFTAG | | | | |
| : ::::: : ::::::: ::::: : ::::::: ::::::: :: : ::::::::::::: | | | | |
| LKGNSHLEIQAYHNLTPEKGWLSSYSVRITWYTEKFWTDTTPDCADVLIHSTYFPCFTAG | | | | |
| 70 | 80 | 90 | 100 | 110 |
| | | | | |
| 130 | 140 | 150 | 160 | 170 |
| EVRRAIRGERLLSCCRFPRAHKHQVPSLQYLAIRVVSHV-RSQGENPTHKQWRRDNRSL | | | | |
| : ::::::: : ::::: : ::::::: ::::: : :: : : :: : :: : | | | | |
| EVRRAIRGEKLLSCCNYPRAHRAQVPSLQFLALVVVQONDPRQRDSTTRKQRRRDYRRGL | | | | |
| 130 | 140 | 150 | 160 | 170 |
| | | | | |
| 180 | 190 | 200 | 210 | |
| RYAKQNSRGDKQRGKPPTEGANFPGLAKVLGILA | | | | |
| : :: : :: :: : : :: : :: : :: : | | | | |
| RLAKQDSRSRKQRSSESPTPRTYFPGVAEVLEILA | | | | |
| 190 | 200 | 210 | | |

FIG. 6

31 / 35

(R.mac
FIG. 7 (versus
 (R.ROD

| | | | | | |
|---|----|----|----|-----|-----|
| ME----ERPPENEGPQREPWDEHVVEVLKELKEEALKHFDPRLTALGNHIYNRHGDITLE | 10 | 20 | 30 | 40 | 50 |
| : | | | | | |
| MAEAPTELPPVDTPLREPGEWIIEILREIKEEALKHFDPRLLIALGKYIYTRHGDITLE | 10 | 20 | 30 | 40 | 50 |
| | 60 | 70 | 80 | 90 | 100 |
| GAGELIRILQRALFIHFRSGCSHSRIGQPGGGNPLSTIPPSRSML | | | | | |
| : | | | | | |
| GARELIKVLQRALFTHFRAGCGHSRIGQTRGGNPLSAIPTPRNMQ | 70 | 80 | 90 | 100 | |

FIG. 7

32/35

FIG.8 (X.mac
 (versus
 (X.ROD

10 20 30 40 50
MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPRELIFQVWORSHEYWHDEQ
::: :::::::::::::::::::: ::::::: :::::::::::::::::::: :::::::
MTOPRETVPNGNSGEETIGEAFAWLNRTVEAINREAVNHLPRELIFQVWQRSHRYWHDEQ
10 20 30 40 50

70 80 90 100 110
GMSOSYTKYRYLCIQQKALFMHCKKGCRCLGEGHGAGGWPGPPPPPPGLA
::: ::::::::::::::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
GMSESYTKYRYLCIIQKAVYMHVRKGCTCLGRGHGPCCWPGPPPPPPGLV
70 80 90 100 110

33 / 35

FIG. 9 (F.mac
versus
F.ROD)

FIG. 9

34 / 35

FIG.10 (TAT.mac
 (versus
 (TAT.ROD

| | | | | |
|---|-----|----|-----|-----|
| 10 | 20 | 30 | 40 | 50 |
| METPLREQENSLESSNERSYISEAAAAIPESANLGEELISOLYRPLEACYNTCYCKKK ::::: :: :: : :: : :: : :: :: :: :: :: :: :: :: :: :: :: :: :: | | | | |
| METPLKAPESSLKSCNEPFSRTSEQDVATQELARQGEEILSOLYRPLEFNNNSCYCKRCC | | | | |
| 10 | 20 | 30 | 40 | 50 |
| YHCQFCFLKKGLGISYEKSHRRRTPKAKANTSSASNERP---IPNRIRLCQPKKAKKE ::::: | | | | |
| YHCQMCFLNKGLGICYERKGRRRTPKKTTHPSPT---POKSISTRGDSOPTKKOKK | | | | |
| 70 | 80 | 90 | 100 | 110 |
| TVEAAVATAPGLGR :::: : : : :: TVEATVETDTGPGR 120 130 | | | | |
| 120 | 130 | | | |

FIG. 10

35 / 35

FIG.11 { ART.mac
 { versus
 { ART.ROD

| | | | | |
|---|------------------------|-------------|-----|-----|
| 10 | 20 | 30 | 40 | 50 |
| MRSHTGEEELRRRLRLIHLHQTSKYGLSWKSAAYRHL | LVDPYPTGSGSANORROKRRRW | | | |
| :: : : :::: ::::: | | :: : : :::: | | |
| MNERADEEGLORKLRLIRLLHOTN----- | PYPOCPGTASORRNRRRW | | | |
| 10 | 20 | | 30 | 40 |
| 70 | 80 | 90 | 100 | 110 |
| RORHQOLLALADRIYSFPDPPTDPLDLAIQQLQNLAIESIPDPPTNIPEALCDLRRIRR | | | | |
| :: : ::::: :: ::::: : :: : :: : :: ::::: :: | | | | |
| KQRWROLALADSIYTDPPPADSPLDQTIQHLQGLTIQELPDPPTHLPESORLAET | | | | |
| 50 | 60 | 70 | 80 | 90 |
| | | | | 100 |

SPQA

FIG. 11

INTERNATIONAL SEARCH REPORT

PCT/FR 88/00025

International Application No.

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) ⁶

According to International Patent Classification (IPC) or to both National Classification and IPC

4 C 07 K 7/10, 7/06, C 12 N 15/00, G 01 N 33/569,
Int. Cl. : A 61 K 39/21, A 61 K 37/02

II. FIELDS SEARCHED

Minimum Documentation Searched ⁷

| Classification System | Classification Symbols |
|-----------------------|--|
| Int. Cl. 4 | A 61 K 39/00, A 61 K 37/00, C 07 K 7/00, G 01 N 33/00, C 12 N 15/00 |

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched ⁸III. DOCUMENTS CONSIDERED TO BE RELEVANT⁹

| Category ¹⁰ | Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹² | Relevant to Claim No. ¹³ |
|------------------------|---|-------------------------------------|
| X, Y | WO, A, 86/02383 (PASTEUR) 24 April 1986, see pages 21-34, 55-60; claims 1-16 -- | 1-36 |
| X, Y | US, A, 4629783 (W.L. COSANT) 16 December 1986, see columns 15,16; claims 1-42 -- | 1-36 |
| X | EP, A, 0199301 (HOFFMANN-LA ROCHE & CO.) 29 October 1986, see columns 27-42; claims 1-42 -- | 1,2,6-12,14- 21,24 |
| X | EP, A, 0187041 (GENENTECH) 9 July 1986, see pages 79-81, claims 1-18; pages 82,83, claims 24-38 -- | 1,2,6-12,14- 21,24 |
| Y | Science, vol. 232, 1985, American Association for the Advancement of Science, (Washington, DC, US) P.J. Kanki et al.: "New human T-lymphotropic retrovirus related to simian T-lymphotropic virus type III (STLV-IIIAGM) pages 238-243 see the whole document -- | 1-36 |
| Y | Science, vol. 233, 18 July 1986, American Association for the Advancement of Science, -- | 1-36 |

¹⁰ Special categories of cited documents:¹⁰ "A" document defining the general state of the art which is not considered to be of particular relevance¹⁰ "E" earlier document but published on or after the international filing date¹⁰ "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)¹⁰ "O" document referring to an oral disclosure, use, exhibition or other means¹⁰ "P" document published prior to the international filing date but later than the priority date claimed¹⁰ "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention¹⁰ "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step¹⁰ "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.¹⁰ "&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search

10 June 1986 (10.06.86)

Date of Mailing of this International Search Report

7 July 1988 (07.07.88)

International Searching Authority

Signature of Authorized Officer

EUROPEAN PATENT OFFICE

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

| Category * | Citation of Document, with indication, where appropriate, of the relevant passages | Relevant to Claim No |
|------------|--|----------------------|
| | (Washington, DC, US) F. Clavel et al.: "Isolation of a new human retrovirus from West African patients with AIDS", pages 343-346 see the whole document -- | |
| Y | Nature, vol. 324, 18/25 December 1986, MacMillan Eds. (Londres, GB) F. Clavel et al.: "Molecular cloning and polymorphism of the human immune deficiency virus type 2", pages 691-695 see the whole document -- | 1-36 |
| Y | Nature, vol. 321, 22 May 1986 MacMillan Eds. (Londres, GB) M. Murphey-Corb et al.: "Isolation of an HTLV-III-related retrovirus from macaques with simian AIDS and its possible origin in asymptomatic mangabeys", pages 435-437 see the whole document -- | 1-36 |
| P,X | WO, A, 87/02038 (ONCOGEN) 9 April 1987, see pages 114-117; claims 31-55 -- | 1,2,6-12,14 21,24 |
| P,X | Journal of Cellular Biochemistry, Supplement 11D, UCLA Symposia on Molecular & Cellular Biology, 29 March - 1 May 1987, Symposium on human retroviruses, 16th Annual Meeting UCLA, see page 44, abstract P112: "Human retroviruses, cancer and AIDS: approaches to prevention and therapy" -- | 1-36 |
| P,Y | FR, A, 2593189 (PASTEUR) 24 July 1987, see page 8, lines 17-26; pages 13-15, claims 1-18 -- | 1-36 |
| P,Y | Nature, vol. 326, No. 6113, 9-15 April 1987, MacMillan Eds. (Londres, GB) H. Kornfeld et al.: "Cloning of HTLV-4 and its relation to simian and human immunodeficiency viruses" pages 610-613, see the whole document -- | 1-36 |
| P,Y | Nature, vol. 326, 16 April 1987, MacMillan Eds. (Londres, GB) M. Guyader et al.: "Genome organization and transactivation of the human immunodeficiency virus type 2, pages 662-669, see the whole document -- | 1-36 |
| P,Y | FEBS Letters, vol. 218, No. 2, June 1987, Eds. Elsevier Siencc Publishers B.V. (Biomedical Division), 1987 Federation of European Biochemical Societies M.J.E. Sternberg et al.: "Prediction of antigenic determinants and secondary structures of the major AIDS virus proteins", pages 231-237 -- | 1-36 |

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

| | | |
|--|------------------------|--|
| | see the whole document | |
|--|------------------------|--|

V. OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE¹

This International search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. Claim numbers because they relate to subject matter not required to be searched by this Authority, namely:

2. Claim numbers because they relate to parts of the International application that do not comply with the prescribed requirements to such an extent that no meaningful International search can be carried out, specifically:

3. Claim numbers....., because they are dependent claims and are not drafted in accordance with the second and third sentences of PCT Rule 8.4(a).

VI. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING²

This International Searching Authority found multiple inventions in this International application as follows:

Claims 1,2,6-21,23-36

Claims 1,2,6-21,23-36 all in part 3-5

Claims 23,25-28,34-36 all in part 22,24,29

1. As all required additional search fees were timely paid by the applicant, this International search report covers all searchable claims of the International application.

2. As only some of the required additional search fees were timely paid by the applicant, this International search report covers only those claims of the International application for which fees were paid, specifically claims:

3. No required additional search fees were timely paid by the applicant. Consequently, this International search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

The additional search fees were accompanied by applicant's protest.

No protest accompanied the payment of additional search fees.

ANNEX TO THE INTERNATIONAL SEARCH REPORT
ON INTERNATIONAL PATENT APPLICATION NO. FR 8800025
SA 20445

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on 23/06/88
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

| Patent document cited in search report | Publication date | Patent family member(s) | | Publication date |
|--|------------------|-------------------------|----------|------------------|
| WO-A- 8602383 | 24-04-86 | FR-A- | 2571968 | 25-04-86 |
| | | AU-A- | 5061785 | 02-05-86 |
| | | EP-A- | 0201540 | 20-11-86 |
| | | JP-T- | 62500592 | 12-03-87 |
| | | WO-A- | 8604336 | 31-07-86 |
| | | AU-A- | 5320086 | 13-08-86 |
| | | EP-A- | 0211022 | 25-02-87 |
| | | JP-T- | 62502095 | 20-08-87 |
| US-A- 4629783 | 16-12-86 | EP-A- | 0201716 | 20-11-86 |
| | | WO-A- | 8606414 | 06-11-86 |
| | | AU-A- | 5572786 | 16-10-86 |
| | | AU-A- | 5773386 | 18-11-86 |
| | | EP-A- | 0220273 | 06-05-87 |
| | | JP-T- | 62502617 | 08-10-87 |
| | | AU-B- | 571128 | 31-03-88 |
| EP-A- 0199301 | 29-10-86 | AU-A- | 5636386 | 23-10-86 |
| | | JP-A- | 62012799 | 21-01-87 |
| EP-A- 0187041 | 09-07-86 | JP-A- | 61233700 | 17-10-86 |
| WO-A- 8702038 | 09-04-87 | AU-A- | 6299286 | 09-04-87 |
| | | BE-A- | 905492 | 25-03-87 |
| | | GB-A- | 2181435 | 23-04-87 |
| | | FR-A- | 2587720 | 27-03-87 |
| | | SE-A- | 8604007 | 26-03-87 |
| | | NL-A- | 8602422 | 16-04-87 |
| | | FR-A- | 2593519 | 31-07-87 |
| | | JP-A- | 63068075 | 26-03-88 |
| FR-A- 2593189 | 24-07-87 | WO-A- | 8704459 | 30-07-87 |
| | | AU-A- | 6891187 | 14-08-87 |
| | | EP-A- | 0239425 | 30-09-87 |

RAPPORT DE RECHERCHE INTERNATIONALE
Demande internationale N° PCT/FR 88/00025

I. CLASSEMENT DE L'INVENTION (si plusieurs symboles de classification sont applicables, les indiquer tous) ¹⁾

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB
 4 C 07 K 7/10, 7/06, C 12 N 15/00, G 01 N 33/569,
 CIB : A 61 K 39/21, A 61 K 37/02

II. DOMAINES SUR LESQUELS LA RECHERCHE A PORTÉ

Documentation minimale consultée ²⁾

| Système de classification | Symboles de classification |
|---------------------------|--|
| CIB ⁴ | A 61 K 39/00, A 61 K 37/00, C 07 K 7/00, G 01 N 33/00, C 12 N 15/00 |

Documentation consultée autre que la documentation minimale dans la mesure où de tels documents font partie des domaines sur lesquels la recherche a porté ³⁾

III. DOCUMENTS CONSIDÉRÉS COMME PERTINENTS ¹⁰⁾

| Catégorie ⁴⁾ | Identification des documents cités, ¹¹⁾ avec indication, si nécessaire, des passages pertinents ¹²⁾ | N° des revendications visées ¹³⁾ |
|-------------------------|--|---|
| X, Y | WO, A, 86/02383 (PASTEUR) 24 avril 1986, voir pages 21-34, 55-60; revendications 1-16 -- | 1-36 |
| X, Y | US, A, 4629783 (W.L. COSANT) 16 décembre 1986, voir colonnes 15,16; revendications 1-42 -- | 1-36 |
| X | EP, A, 0199301 (HOFFMANN-LA ROCHE & CO.) 29 octobre 1986, voir colonnes 27-42; revendications 1-42 -- | 1,2,6-12, 14-21,24 |
| X | EP, A, 0187041 (GENENTECH) 9 juillet 1986, voir pages 79-81, revendications 1-18; pages 82,83, revendications 24-38 -- | 1,2,6-12, 14-21,24 |
| Y | Science, vol. 232, 1985, American Association for the Advancement of Science, (Washington, DC, US) P.J. Kanki et al.: "New human T-lymphotropic retrovirus related to simian T-lymphotropic virus type III (STLV-IIIAGM)" | 1-36 |

* Catégories spéciales de documents cités: ¹¹⁾

- « A » document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- « E » document antérieur, mais publié à la date de dépôt international ou après cette date
- « L » document pouvant poser un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- « Q » document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- « P » document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- « T » document ultérieur publié postérieurement à la date de dépôt international ou à la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- « X » document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive
- « Y » document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier.
- « & » document qui fait partie de la même famille de brevets

IV. CERTIFICATION

Date à laquelle la recherche internationale a été effectivement
achevée
10 juin 1986

Date d'expédition du présent rapport de recherche internationale

- 7 JUL 1988

Administration chargée de la recherche internationale
OFFICE EUROPÉEN DES BREVETS

Signature du fonctionnaire autorisé

P.C.G. VAN DER PUTTEN

| III. DOCUMENTS CONSIDÉRÉS COMME PERTINENTS | | (SUITE DES RENSEIGNEMENTS INDIQUÉS SUR LA DEUXIÈME FEUILLE) |
|--|---|---|
| Catégorie* | Identification des documents cités, avec indication, si nécessaire, des passages pertinents | N° des revendications visées |
| | pages 238-243 voir le document en entier -- | |
| Y | Science, vol. 233, 18 juillet 1986, American Association for the Advancement of Science, (Washington, DC, US) F. Clavel et al.: "Isolation of a new human retrovirus from West African patients with AIDS", pages 343-346 voir le document en entier -- | 1-36 |
| Y | Nature, vol. 324, 18/25 décembre 1986, MacMillan Eds. (Londres, GB) F. Clavel et al.: "Molecular cloning and polymorphism of the human immune deficiency virus type 2", pages 691-695 voir le document en entier -- | 1-36 |
| Y | Nature, vol. 321, 22 mai 1986, MacMillan Eds. (Londres, GB) M. Murphey-Corb et al.: "Isolation of an HTLV-III-related retrovirus from macaques with simian AIDS and its possible origin in asymptomatic mangabeys", pages 435-437 voir le document en entier -- | 1-36 |
| P,X | WO, A, 87/02038 (ONCOGEN) 9 avril 1987, voir apges 114-117; revendicaions 31-55 -- | 1,2,6-12, 14-21,24 |
| P,X | Journal of Cellular Biochemistry, Supplement 11D, Ucla Symposia on Molecular & Cellular Biology, 29 mars - 1 mai 1987, Symposium on human retroviruses, 16th Annual Meeting UCLA, voir page 44, abrégé P112: "Human retroviruses, cancer and AIDS: approaches to prevention and therapy" -- | 1-36 |
| P,Y | FR, A, 2593189 (PASTEUR) 24 juillet 1987, voir page 8, lignes 17-26; pages 13-15, revendications 1-18 -- | 1-36 |
| P,Y | Nature, vol. 326, no. 6113, 9-15 avril 1987, MacMillan Eds.(Londres, GB) H. Kornfeld et al.: "Cloning of HTLV-4 and its relation to simian and human immunodeficiency viruses" pages 610-613, voir le document en entier -- | 1-36 |

| III. DOCUMENTS CONSIDÉRÉS COMME PERTINENTS | | (SUITE DES RENSEIGNEMENTS INDICUÉS SUR LA DEUXIÈME FEUILLE) |
|--|---|---|
| Catégorie* | Identification des documents cités, avec indication, si nécessaire, des passages pertinents | N° des revendications visées |
| P,Y | Nature, vol. 326, 16 avril 1987, MacMillan Eds. (Londres, GB) M. Guyader et al.: "Genome organization and transactivation of the human immuno- deficiency virus type 2, pages 662-669, voir le document en entier | 1-36 |
| P,Y | FEBS Letters, vol. 218, no. 2, juin 1987, Eds. Elsevier Siencc Publishers B.V. (Biomedical Division), 1987 Federation of European Biochemical Societies M.J.E. Sternberg et al.: "Prediction of antigenic determinants and secondary structures of the major AIDS virus proteins", pages 231-237 voir le document en entier | 1-36 |
| | ----- | |

SUITE DES RENSEIGNEMENTS INDICÉS SUR LA DEUXIÈME FEUILLE

**ANNEXE AU RAPPORT DE RECHERCHE INTERNATIONALE
RELATIF A LA DEMANDE INTERNATIONALE NO.**

FR 8800025
SA 20445

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche international visé ci-dessus.

Lesdits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du 23/06/88

Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets.

| Document brevet cité au rapport de recherche | Date de publication | Membre(s) de la famille de brevet(s) | Date de publication |
|---|------------------------|--|--|
| WO-A- 8602383 | 24-04-86 | FR-A- 2571968 AU-A- 5061785 EP-A- 0201540 JP-T- 62500592 WO-A- 8604336 AU-A- 5320086 EP-A- 0211022 JP-T- 62502095 | 25-04-86 02-05-86 20-11-86 12-03-87 31-07-86 13-08-86 25-02-87 20-08-87 |
| US-A- 4629783 | 16-12-86 | EP-A- 0201716 WO-A- 8606414 AU-A- 5572786 AU-A- 5773386 EP-A- 0220273 JP-T- 62502617 AU-B- 571128 | 20-11-86 06-11-86 16-10-86 18-11-86 06-05-87 08-10-87 31-03-88 |
| EP-A- 0199301 | 29-10-86 | AU-A- 5636386 JP-A- 62012799 | 23-10-86 21-01-87 |
| EP-A- 0187041 | 09-07-86 | JP-A- 61233700 | 17-10-86 |
| WO-A- 8702038 | 09-04-87 | AU-A- 6299286 BE-A- 905492 GB-A- 2181435 FR-A- 2587720 SE-A- 8604007 NL-A- 8602422 FR-A- 2593519 JP-A- 63068075 | 09-04-87 25-03-87 23-04-87 27-03-87 26-03-87 16-04-87 31-07-87 26-03-88 |
| FR-A- 2593189 | 24-07-87 | WO-A- 8704459 AU-A- 6891187 EP-A- 0239425 | 30-07-87 14-08-87 30-09-87 |